



University  
of Glasgow

Ferreira de Lima, Danilo Enoque (2014) *Top-antitop cross section measurement as a function of the jet multiplicity in the final state and beyond the Standard Model top-antitop resonances search at the ATLAS detector at CERN*. PhD thesis.

<http://theses.gla.ac.uk/5015/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given



PH. D. THESIS

---

Top-antitop cross section measurement  
as a function of the jet multiplicity in  
the final state and beyond the Standard  
Model top-antitop resonances search at  
the ATLAS detector at CERN

---

© Danilo Enoque Ferreira de Lima

School of Physics and Astronomy  
College of Science and Engineering

Submitted in fulfilment of the requirements for the degree  
of Doctor in Philosophy at the University of Glasgow

February 2014

## Abstract

The top quark is the heaviest particle in the Standard Model, with a strong coupling to the Higgs boson. It is often seen as a window to new physics, therefore understanding its production is a key ingredient for testing the Standard Model or physics Beyond the Standard Model. In this document, the production cross section of top-antitop pairs in its semileptonic decay channel is measured as a function of the jet multiplicity in the ATLAS experiment, using proton-proton collisions at the center-of-mass energy of  $\sqrt{s} = 7$  TeV. The top-antitop production with extra jets is the main background for many analyses, including the top-antitop-Higgs production studies. The analysis performed is extended in a search for Beyond the Standard Model physics which predicts a resonance decaying in a top-antitop pair, using ATLAS data at center-of-mass energy of  $\sqrt{s} = 7$  TeV. The latter analysis is repeated for ATLAS data collected with  $\sqrt{s} = 8$  TeV. Performance studies of  $b$ -tagging algorithms in the ATLAS Trigger System are also presented.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>I</b>	<b>Theoretical foundations</b>	<b>5</b>
<b>2</b>	<b>Theory overview</b>	<b>6</b>
2.1	The Standard Model . . . . .	6
2.1.1	Matter fields and electroweak interactions . . . . .	9
2.1.2	Quantum Chromodynamics . . . . .	11
2.1.3	Electroweak symmetry breaking mechanism . . . . .	13
2.2	The Standard Model and the top quark . . . . .	14
2.3	Top-antitop pair generation at the LHC . . . . .	16
2.4	Monte Carlo event generators . . . . .	18
2.4.1	Factorisation theorem and perturbative treatment . . . . .	20
2.4.2	Parton showers . . . . .	22
2.4.3	Next-to-leading order matrix element generators . . . . .	26
2.4.4	Hadronisation . . . . .	27
2.4.5	Underlying events . . . . .	28
2.5	Beyond the Standard Model . . . . .	29
2.6	Summary . . . . .	30
<b>II</b>	<b>The experimental setup</b>	<b>31</b>
<b>3</b>	<b>The ATLAS experiment</b>	<b>32</b>
3.1	The ATLAS detector . . . . .	32
3.1.1	Inner Detector . . . . .	35
3.1.2	Calorimeters . . . . .	36
3.1.3	Muon Spectrometer . . . . .	37
3.1.4	The ATLAS Trigger System . . . . .	38

3.2	Multiple interactions in ATLAS . . . . .	41
3.3	Electron reconstruction and identification . . . . .	41
3.4	Muon reconstruction . . . . .	47
3.5	Jet algorithms . . . . .	51
3.6	$b$ -tagging algorithms . . . . .	54
3.7	Missing transverse energy reconstruction . . . . .	56
3.8	Summary . . . . .	58
<b>4</b>	<b><math>b</math>-jet trigger performance studies</b>	<b>59</b>
4.1	$b$ -jet trigger chains configuration . . . . .	59
4.2	$b$ -tagging algorithms . . . . .	61
4.3	Data to Monte Carlo simulation comparison of with $\sqrt{s} = 7$ TeV data . . . . .	63
4.4	Data to simulation comparison of the $b$ -jet combined “physics” trigger using $\sqrt{s} = 8$ TeV, 2012 data . . . . .	65
4.5	Data to simulation comparison in heavy flavour enriched sample with ATLAS 2012 $\sqrt{s} = 8$ TeV data . . . . .	68
4.6	Summary . . . . .	72

### III Physics Analyses 73

<b>5</b>	<b>Top-antitop differential cross section measurement as a func- tion of the jet multiplicity in the final state</b>	<b>74</b>
5.1	Motivation . . . . .	74
5.2	Top-antitop signal simulation and background estimates . . . . .	75
5.3	Top-antitop event selection . . . . .	78
5.3.1	Trigger and pile up-related selection . . . . .	78
5.3.2	Lepton selection . . . . .	79
5.3.3	Jet selection . . . . .	80
5.3.4	Missing energy requirements . . . . .	80
5.4	Data-driven W+jets background estimate . . . . .	81
5.5	Data-driven QCD multi-jets background estimate . . . . .	83
5.6	Corrections applied in simulation . . . . .	86
5.7	Data to signal and background comparison . . . . .	89
5.8	Systematic uncertainties estimate at reconstruction level . . . . .	92
5.9	Unfolding the effect of the detector . . . . .	99
5.10	Propagation of systematic uncertainties through the unfolding procedure . . . . .	105

5.11	Results at particle level and discussion . . . . .	107
5.12	Correction factors and consistency checks for selections with jet cuts at 40 GeV, 60 GeV and 80 GeV . . . . .	131
<b>6</b>	<b>Top-antitop resonances search at <math>\sqrt{s} = 7</math> TeV</b>	<b>147</b>
6.1	Benchmark models and motivation . . . . .	148
6.2	Search strategy . . . . .	148
6.3	Background modelling . . . . .	150
6.4	Event selection . . . . .	152
6.5	Corrections applied to simulation and data . . . . .	156
6.6	Event reconstruction . . . . .	157
6.7	Systematic uncertainties . . . . .	160
6.8	Data to expectation comparison . . . . .	163
6.9	Limit setting and summary . . . . .	171
<b>7</b>	<b>Top-antitop resonances search at <math>\sqrt{s} = 8</math> TeV</b>	<b>177</b>
7.1	Differences with respect to the $\sqrt{s} = 7$ TeV analysis . . . . .	177
7.2	Multi-jet background modelling . . . . .	178
7.3	Event reconstruction and results . . . . .	191
7.4	Limit setting and summary . . . . .	198
<b>8</b>	<b>Summary</b>	<b>204</b>
<b>A</b>	<b>Top-antitop + jets control plot distributions</b>	<b>206</b>

# List of Tables

2.1	Properties of the fundamental particles of the Standard Model. Information extracted from [7]. Particle's masses were rounded to show their order of magnitude. Latest measurements including errors can be found in [7]. . . . .	7
5.1	Event yields for data and MC simulation in the electron and muon channels, selected with a 25 GeV jet $p_T$ threshold. The number of events passing all selection requirements are shown as a function of the reconstructed jet multiplicity ( $n_{\text{jets}}^{\text{reco}}$ ). Alpgen+Herwig is used for the $t\bar{t}$ simulation and MC expectations are normalised to an integrated luminosity of $4.7 \text{ fb}^{-1}$ . The uncertainties on the expected values include systematic uncertainties. . . . .	91
5.2	Uncertainties on event yields at reconstruction level in the electron channel, selected with a 25 GeV jet $p_T$ threshold. Alpgen is used for the $t\bar{t}$ simulation. The uncertainties are shown as a percentage of the expected $t\bar{t}$ signal. . . . .	113
5.3	Uncertainties on event yields at reconstruction level in the muon channel, selected with a 25 GeV jet $p_T$ threshold. Alpgen is used for the $t\bar{t}$ simulation. The uncertainties are shown as a percentage of the expected $t\bar{t}$ signal. . . . .	114
5.4	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The $p_T$ cut on the jets is 25 GeV. . . . .	115
5.5	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The $p_T$ cut on the jets is 25 GeV. . . . .	116
5.6	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The $p_T$ cut on the jets is 40 GeV. . . . .	141

5.7	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The $p_T$ cut on the jets is 40 GeV. . .	142
5.8	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The $p_T$ cut on the jets is 60 GeV. .	143
5.9	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The $p_T$ cut on the jets is 60 GeV. . .	144
5.10	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The $p_T$ cut on the jets is 80 GeV. .	145
5.11	Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The $p_T$ cut on the jets is 80 GeV. . .	146
6.1	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 7\text{TeV}$ in the resolved electron channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds. . . . .	164
6.2	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 7\text{TeV}$ in the resolved muon channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds. . . . .	170
6.3	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 7\text{TeV}$ in the boosted electron channel with statistical uncertainties for data and background samples, followed by the systematic uncertainty for all background samples. . . . .	170
6.4	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 7\text{TeV}$ in the boosted muon channel with the statistical uncertainties for data and background samples, followed by the systematic uncertainty for the background samples. . . . .	170



6.5	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the resolved electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	171
6.6	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the resolved muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	174
6.7	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 7\text{TeV}$ , in the boosted electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	175
6.8	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the boosted muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	176
7.1	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 8\text{TeV}$ in the resolved electron channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds. . . . .	191
7.2	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 8\text{TeV}$ in the resolved muon channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds. . . . .	191
7.3	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 8\text{TeV}$ in the boosted electron channel with statistical uncertainties for data and background samples, followed by the total systematic uncertainty for the background samples. . . . .	192

7.4	Total contribution of each of the background samples in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 8\text{TeV}$ in the boosted muon channel with the statistical uncertainties for data and background samples, followed by the total systematic uncertainty for the background samples. . . . .	192
7.5	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the resolved electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	200
7.6	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the resolved muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	201
7.7	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis at $\sqrt{s} = 8\text{TeV}$ , in the boosted electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	202
7.8	Systematic uncertainties from all backgrounds in percentage variation of the $t\bar{t}$ sample, in the $t\bar{t}$ resonances analysis, in the boosted muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied). . . . .	203

# List of Figures

2.1	Top-antitop main production diagrams at the LHC, at tree-level.	17
2.2	Top quark decays . . . . .	17
2.3	Single top production diagrams at tree-level. . . . .	18
2.4	Simplified schematic view of simulation steps necessary for physics analyses. . . . .	19
2.5	Graphical representation of the Factorisation Theorem, Equation 2.31. The symbols are the same as in Equation 2.31, except that $f_1, f_2, \dots, f_n$ represent multiple particles in the final state $f$ . . . . .	20
2.6	Simplified schematic of a parton splitting in which a parton $a$ emits a parton $b$ and proceeds as $b'$ . . . . .	23
2.7	$t\bar{t}$ next-to-leading-order diagram examples. . . . .	26
3.1	Schematic view of the Large Hadron Collider and other particle accelerators with the indication for the experiments built in the LHC ring. All credits to ©CERN. . . . .	33
3.2	A schematic view of the ATLAS experiment. All credits to ©CERN. . . . .	34
3.3	A simplified schematic view of the ATLAS Trigger System. Extracted from [77]. . . . .	39
3.4	Simplified schematic that shows the structure of the trigger chains.	42
3.5	Schematic representation of the Trigger Towers used to calculate the electron/photon-related Level 1 trigger threshold sums. The core of $2 \times 2$ trigger towers in the electromagnetic calorimeter is required to contain the sum of two Trigger Towers horizontally or vertically that satisfy the minimum threshold. Isolation veto using the ring of cells around the center ones and the hadronic calorimeter energy sums can also be implemented in some chains. Extracted from [11]. . . . .	43

3.6	Schematic representation of the Trigger Tower sum configuration for the jet-related triggers at Level 1. Extracted from [11]. The jet trigger algorithms are based on jet elements which have the size of $2 \times 2$ Trigger Towers. The Region of Interest is shaded. For scans using $6 \times 6$ windows, there are four possible windows containing a Region of Interest, but in the $8 \times 8$ case, the Region of Interest is required to be in the center position, to avoid the possibility of two jets in a single window. . . . .	43
3.7	The mean number of proton-proton interactions per bunch crossing in ATLAS is shown for the data taking in 2011 (left) and 2012 (right). For 2011, the set up after the Technical Stop in September (with $\beta^* = 1.0$ m) is shown in red and the set up before it is shown in blue (with $\beta^* = 1.5$ m). ATLAS performance public plot not produced by the author. More information about the measurement can be found in [78]. Entries in $\langle \mu \rangle \sim 0$ arise from pilot bunches that were present in many early LHC fills. . . . .	44
3.8	Results of the measurement of the electron energy scale in $Z \rightarrow e^+e^-$ decays and in $J/\Psi \rightarrow e^+e^-$ decays in ATLAS 2010 data, for $ \eta  < 0.6$ (left) and $1.53 <  \eta  < 1.8$ (right). Extracted from [79]. . . . .	45
3.9	Fit of data and simulation for the electron energy resolution estimate from $J/\Psi \rightarrow e^+e^-$ decays using ATLAS 2010 data. Extracted from [79]. . . . .	46
3.10	Efficiency measurement results using the Tag And Probe method in $Z \rightarrow e^+e^-$ decays in 2010 ATLAS data for the electron identification (left) and the electron reconstruction efficiencies. Extracted from [79]. . . . .	47
3.11	Sum in quadrature of the Muon Spectrometer and the Inner Detector muon resolutions as a function of the transverse momentum in four pseudo-rapidity regions using $W \rightarrow \mu\nu$ events in ATLAS 2010 data. This is the result of a preliminary analysis, on which there were shortcomings in the simulation of intrinsic resolution and module misalignment [81]. Extracted from [81]. . .	49

3.12	Muon reconstruction efficiency not considering the isolation requirement, measured using $Z$ -boson decays into pairs of muons. In the left figure, the Inner Detector reconstruction efficiency is shown. In the right figure, the efficiency of reconstructing Combined Muons, relative to the Inner Detector efficiency is shown. This was done with 2010 ATLAS data and it was extracted from [80]. . . . .	51
3.13	Ratio of the jet energy scale in data and simulation for Anti- $k_t$ $R = 0.4$ jets built using the EM scale (left) or using the LCW method (right) for 2011 ATLAS data. Extracted from [87]. . . .	53
3.14	The efficiency in data and simulation (left) and their ratio (right) for the MV1 b-tagging algorithm in its 70% efficiency working point, calculated using ATLAS 2011 data and the $p_T^{\text{rel}}$ method [90, 92]. Extracted from [92]. . . . .	56
3.15	Left: resolution of the missing transverse energy measured in 2010 ATLAS $\sqrt{s} = 7\text{TeV}$ data with the respective fits in each channel. Right: uncertainty in the missing transverse energy scale from Monte Carlo simulation of $W$ -boson decays into an electron and a neutrino. Extracted from [93]. . . . .	58
4.1	Simplified representation of the $B$ -hadron decay in a secondary vertex. Extracted from [77]. . . . .	61
4.2	The signed $S(d_0)$ for the selected tracks at the Event Filter, using 2011 ATLAS data. . . . .	64
4.3	The track multiplicity for selected tracks at the Event Filter, using 2011 ATLAS data. . . . .	64
4.4	The tracks' transverse momenta for selected tracks at the Event Filter, using 2011 ATLAS data. . . . .	65
4.5	Combined tagger weight for physics trigger using the impact parameter significance and the secondary vertex likelihood-based taggers, calculated from Level 2 and Event Filter tracks in low $p_T$ jets identified by the Level 1. . . . .	66
4.6	IP3D tagger for physics trigger using the transverse and longitudinal impact parameter significances, calculated using Level 2 and Event Filter tracks in low $p_T$ jets identified by the Level 1. . . . .	67
4.7	Data to simulation comparison for the physics trigger with flavour association for the Level 2 SV variables. . . . .	67

4.8	Data to simulation comparison for the physics trigger with flavour association for the Event Filter SV variables. . . . .	68
4.9	Data to simulation comparison for the physics trigger with flavour association for the mass of the SV. . . . .	68
4.10	Combined tagger weight using the impact parameter significance and the secondary vertex likelihood-based taggers, calculated from Level 2 and Event Filter tracks in low $p_T$ jets identified by the Level 1. . . . .	70
4.11	Data to simulation comparison for the calibration trigger with flavour association for the IP3D tagger. . . . .	70
4.12	Data to simulation comparison for the calibration trigger with flavour association for the Level 2 SV variables. . . . .	71
4.13	Data to simulation comparison for the calibration trigger with flavour association for the Event Filter SV variables. . . . .	71
4.14	Data to simulation comparison for the calibration trigger with flavour association for the mass of the SV. . . . .	72
5.1	Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the $t\bar{t}$ signal ( $p_T > 25$ GeV). . . . .	90
5.2	Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the $t\bar{t}$ signal ( $p_T > 40$ GeV). . . . .	93
5.3	Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the $t\bar{t}$ signal ( $p_T > 60$ GeV). . . . .	94
5.4	Jet multiplicity in the electron (left) and muon (right) channel using Alpgen simulation for the $t\bar{t}$ signal ( $p_T > 80$ GeV). . . . .	95
5.5	Jet multiplicity in the electron (left) and muon (right) channels using Alpgen+Herwig simulation for the $t\bar{t}$ signal with different Anti- $k_t$ jet transverse momentum cuts applied. In this figure, for the 60GeV plot, the 7 jet bin represents events with $\geq 7$ jets, and in the 80GeV plot, the 6 jet bin represents events with $\geq 6$ jets. . . . .	96
5.6	The $1 - f'_{fakes}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	102
5.7	The $1 - f_{np3}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	103

5.8	The migration matrix using the Alpgen $t\bar{t}$ signal sample with a selection using a jet $p_T$ cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	103
5.9	The $f_{reco}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	103
5.10	The closure test using the Alpgen $t\bar{t}$ signal sample for input and corrections with a jet $p_T$ cut at 25 GeV for the selection. The results for the electron (left) and muon (right) channels are shown. . . . .	104
5.11	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 25 GeV. . . . .	109
5.12	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 40 GeV. . . . .	110
5.13	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 60 GeV. . . . .	111
5.14	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 80 GeV. . . . .	112
5.15	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 25 GeV. . . . .	117

5.16	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 40 GeV. . . . .	118
5.17	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 60 GeV. . . . .	119
5.18	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 80 GeV. . . . .	120
5.19	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 25 GeV. . . . .	122
5.20	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 40 GeV. . . . .	123
5.21	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 60 GeV. . . . .	124
5.22	The unfolded data using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 80 GeV. . . . .	125



5.23	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 25 GeV. . . .	126
5.24	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 40 GeV. . . .	127
5.25	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 60 GeV. . . .	128
5.26	The unfolded cross section using the Alpgen $t\bar{t}$ signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The $p_T$ cut on the jets is 80 GeV. . . .	129
5.27	Jet gap fraction for $ y  < 0.8$ , extracted from [104]. . . . .	130
5.28	The closure test using the Alpgen $t\bar{t}$ signal sample for input and corrections with a jet $p_T$ cut at 40 GeV for the selection. The results for the electron (left) and muon (right) channels are shown. . . . .	132
5.29	The $1 - f'_{fakes}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	133
5.30	The $1 - f_{np3}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	133
5.31	The migration matrix using the Alpgen $t\bar{t}$ signal sample with a selection using a jet $p_T$ cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	133
5.32	The $f_{reco}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	134

5.33	The closure test using the Alpgen $t\bar{t}$ signal sample for input and corrections with a jet $p_T$ cut at 60 GeV for the selection. The results for the electron (left) and muon (right) channels are shown. . . . .	135
5.34	The $1 - f'_{fakes}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	136
5.35	The $1 - f_{np3}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	136
5.36	The migration matrix using the Alpgen $t\bar{t}$ signal sample with a selection using a jet $p_T$ cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	136
5.37	The $f_{reco}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	137
5.38	The closure test using the Alpgen $t\bar{t}$ signal sample for input and corrections with a jet $p_T$ cut at 80 GeV for the selection. The results for the electron (left) and muon (right) channels are shown. . . . .	138
5.39	The $1 - f'_{fakes}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	139
5.40	The $1 - f_{np3}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	139
5.41	The migration matrix using the Alpgen $t\bar{t}$ signal sample with a selection using a jet $p_T$ cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	139
5.42	The $f_{reco}$ correction using the Alpgen $t\bar{t}$ signal sample with a jet $p_T$ cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown. . . . .	140
6.1	Leading jet transverse momentum in the resolved selection. . . .	165
6.2	Leading jet transverse momentum in the boosted selection. . . .	165
6.3	Reconstructed mass of the leptonically decaying top quark in the boosted selection. . . . .	165

6.4	Mass of the hadronically decaying top quark in the boosted selection, reconstructed by the mass of the large- $R$ jet, with no requirement that the mass of the large- $R$ jet is greater than 100GeV. . . . .	166
6.5	Last splitting scale for the large- $R$ jet in the boosted selection, $\sqrt{d_{12}}$ , without the cut in this variable, in this plot. . . . .	166
6.6	Reconstructed invariant mass of the $t\bar{t}$ system for selected events in the resolved scenario. . . . .	167
6.7	Reconstructed invariant mass of the $t\bar{t}$ system for selected events in the boosted scenario. . . . .	168
6.8	Reconstructed invariant mass of the $t\bar{t}$ system for selected events in both resolved and boosted topologies and both electron and muon channels added in a single hisogram. The $Z'$ signal with invariant mass of 1.6TeV and the Kaluza-Klein gluon with an invariant mass of 2.0TeV are overlayed in this plot, with their cross section multiplied by ten to make the effect visible. . . .	169
6.9	Observed and expected upper cross section times branching ratio limit for a narrow $Z'$ resonance. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [105].	173
6.10	Observed and expected upper cross section times branching ratio limit for a Kaluza-Klein gluon. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [105].	173
7.1	$\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the electron (left) and muon (right) channels, for the resolved selection. . . . .	180
7.2	$\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the muon channel, for the boosted selection. In the muon channel, to reduce the statistical uncertainty, this parametrisation is only used for muons with min ( $\Delta R(\text{lepton}, \text{jet})$ ) $\leq 0.4$ and a parametrisation solely described by the muon $p_T$ is used otherwise. . . . .	181

7.3	$\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ for the electron (left) and muon (right) channels, in the boosted selection, which is used if the $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . In the muon channel, the previous criteria might not be satisfied and a parametrisation in function of both these variables is used in such a case. . . . .	181
7.4	The number of $b$ -tagged events over all events in the Control Region. For these plots, no $S(d_0)$ and $b$ -tagging cut were required for all events. The loose criteria is required. . . . .	183
7.5	The fraction of $b$ -tagged jets versus the $ S(d_0) $ of the event in the Control Region. For these plots, no $S(d_0)$ and $b$ -tagging cut were required for all events. The loose criteria is required. . . . .	184
7.6	$\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the electron (left) and muon (right) channels, in the resolved selection, only for $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . . . . .	185
7.7	$\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, only for $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . . . . .	185
7.8	$\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the muon channel, in the resolved selection (left) and boosted selection (right), only for $\min(\Delta R(\text{lepton}, \text{jet})) \leq 0.4$ . . . . .	185
7.9	Systematic uncertainty in $\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ and the $\min(\Delta R(\text{lepton}, \text{jet}))$ in the electron (left) and muon (right) channels, for the resolved selection. . . . .	186
7.10	Systematic uncertainty in $\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ and the $\min(\Delta R(\text{lepton}, \text{jet}))$ in the muon channel, for the boosted selection. In the muon channel, to reduce the statistical uncertainty, this parametrisation is only used for muons with $\min(\Delta R(\text{lepton}, \text{jet})) \leq 0.4$ and a parametrisation solely described by the muon $p_T$ is used otherwise. . . . .	186
7.11	Systematic uncertainty in $\epsilon_{\text{eff}}$ parametrised as a function of the lepton $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, which is used if the $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . In the muon channel, the previous criteria might not be satisfied and a parametrisation as a function of both these variables is used in such a case. . . . .	187

7.12	Systematic uncertainty in $\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the electron (left) and muon (right) channels, in the resolved selection, only for $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . . . . .	187
7.13	Systematic uncertainty in $\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, only for $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ . . . . .	187
7.14	Systematic uncertainty in $\epsilon_{\text{fake}}$ parametrised as a function of the lepton $p_T$ and the closest jet to lepton $p_T$ , for the muon channel, in the resolved selection (left) and boosted selection (right), only for $\min(\Delta R(\text{lepton}, \text{jet})) \leq 0.4$ . . . . .	188
7.15	$m_{t\bar{t}}$ variable calculated in the resolved scenario, in the QCD multi-jets enriched control region, for the electron (left) and muon (right) channels. . . . .	189
7.16	$m_{t\bar{t}}$ variable calculated in the boosted scenario, in the QCD multi-jets enriched control region, for the electron (left) and muon (right) channels. . . . .	190
7.17	Transverse momentum of the leading jet in the resolved scenario.	193
7.18	Transverse momentum of the large- $R$ jet chosen as the hadronically decaying top quark candidate in the boosted selection. . .	193
7.19	Invariant mass of the leptonically decaying top quark candidate in the boosted selection. . . . .	194
7.20	Mass of the large- $R$ jet chosen as the hadronically decaying top quark candidate in the boosted selection. . . . .	194
7.21	First splitting scale, $\sqrt{d_{12}}$ for the large- $R$ jet chosen as the hadronically decaying top quark candidate in the boosted selection. . . . .	194
7.22	Reconstructed invariant mass of the $t\bar{t}$ system in the resolved scenario. . . . .	195
7.23	Reconstructed invariant mass of the $t\bar{t}$ system in the boosted scenario. . . . .	196
7.24	Reconstructed invariant mass of the $t\bar{t}$ system for the resolved, boosted, electron and muon channels summed in a single histogram. One mass point for each benchmark model in the analysis is overlaid with the background, having their production cross section multiplied by five. . . . .	197

7.25	Observed and expected upper cross section times branching ratio limit for a narrow $Z'$ resonance. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [115].	198
7.26	Observed and expected upper cross section times branching ratio limit for a Kaluza-Klein gluon. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [115].	199
A.1	Data to expected signal and background comparison of all jets $p_T$ from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44, 48, 49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	207
A.2	Data to expected signal and background comparison of the highest transverse momentum jet $p_T$ from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44, 48, 49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	208
A.3	Data to expected signal and background comparison of the second highest transverse momentum jet $p_T$ from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44, 48, 49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	209

A.4	Data to expected signal and background comparison of the third highest transverse momentum jet $p_T$ from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44,48,49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	210
A.5	Data to expected signal and background comparison of the fourth highest transverse momentum jet $p_T$ from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44,48,49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	211
A.6	Data to expected signal and background comparison of the lepton transverse momentum from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44,48,49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	212
A.7	Data to expected signal and background comparison of the lepton pseudo-rapidity from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44,48,49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	213
A.8	Data to expected signal and background comparison of the missing transverse energy from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25GeV. The Alpgen+Herwig [44,48,49] $t\bar{t}$ MC sample was used within the data driven and MC predictions. . . . .	214

# Acknowledgements

This thesis could not have been written without the careful support and the interesting discussions provided by my supervisors, Dr. Craig Buttar and Prof. Dr. Anthony Doyle. Their advices were complemented by the constant guidance of Dr. James Ferrando and Dr. Sarah Allwood-Spiers. The discussions with Dr. Peter Bussey were also essential for the implementation of the top-antitop jet multiplicity unfolding method. The discussions and collaboration of Dr. Cristina Oropeza were almost as important as the friendly support she, Ignacio Santiago and Flavia Velásquez gave me, which kept me sane in the most anxious times. My friends, Felipe Martins, André Mendes, Gabriela Roméro, Marcelo Domingues, Marcelo Larcher, Ramón Aguilera, Lino Ferraz, Isabela Salgado and Luma Miranda are not to be forgotten, as they kept me focused even when separated by an ocean. In parallel with my oldest friends, Francesca Minelli had an important role in both supporting me and distracting me, when necessary. The encouragement given by Dr. Denis Damazio, Dr. José de Seixas and Dr. Arthur Moraes was very important to allow me to even think of starting this long enterprise. Last, but not least, I thank my whole family that did not spare efforts to keep me going in the hardest of times, particularly my mother, Angela, my father, Enoque, and my brother, Diogo.



# Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own research work in the Experimental Particle Physics group of the School of Physics and Astronomy in the University of Glasgow. It has not been submitted for any other degree at the University of Glasgow or any other institution.

Danilo Enoque Ferreira de Lima

# Author's Contribution

The research detailed in this document is the result of a collaborative effort in the ATLAS experiment. In parallel, a clear description of the work requires an account of all sectors involved in the analyses, even the ones in which the author did not contribute directly. The list below is presented to clarify the author's contribution in each chapter.

- Chapter 4: The author contributed the data to simulation comparison plots, which were used in the ATLAS Collaboration, to perform a recalibration of the  $b$ -jet trigger taggers.
- Chapter 5: The author performed the data to simulation comparison at reconstruction level, including a calculation of all systematic uncertainty effects, all the Monte Carlo simulation backgrounds and the implementation of the correction factors (but not including the data-driven background parametrisation estimates for the  $W + \text{jets}$  and QCD multi-jets backgrounds); the full unfolding procedure described, calculating all the correction factors from simulation; the propagation of systematic uncertainties through the unfolding procedure; the final unfolded data to particle-level simulation comparison.
- Chapter 6: The author contributed the data to simulation comparison estimates, including all systematic effect estimates and all Monte Carlo simulation background estimates (but not the  $W + \text{jets}$  and QCD multi-jets backgrounds' parametrisation estimate).
- Chapter 7: The author contributed the data to simulation comparison estimates, including all systematic effects estimates and all Monte Carlo simulation background estimates (but not the  $W + \text{jets}$  data-driven parametrisation estimate). The author also contributed in the QCD multi-jets background parametrisation and estimation in this analysis.

# Chapter 1

## Introduction

Particle physics is a recent topic in the history of science, although the idea of dividing matter in elementary building blocks is as old as Democritus' (460 BC - 370 BC) atomic hypothesis [1] <sup>1</sup>. The ancient view of matter and how it interacts has been the theme of many discussions in the history of mankind, evolving from the classical Greek philosophers to the modern view of atomic structure. The idea of indivisible fundamental elements of matter has been extended in the 20<sup>th</sup> century, to include the experimental evidence on the structure of the atom, which led to the development of Quantum Mechanics [3] and, later, Quantum Field Theory [4], which are widely accepted. The proton and neutron in the atom were then subdivided, in this view of matter, into elementary constituents rich in the way they behave and in implications for the future of physics.

Questions could be asked on whether the fundamental elements of matter do exist or whether they are a mathematical tool to describe the observed phenomena, which would lead us to question what it means to observe something. This point will not be discussed in this document, since our goal will be simply to compare the experimental results with the theoretical predictions. Observed phenomena is understood, in this text, as any direct or indirect result of a physical experiment that can be perceived through any rational being's senses, which allows this being to reach a conclusion that the experiment is the most probable cause of the observed phenomena. In the context of Quantum Mechanics, the predictions are made in terms of probabilities, therefore, the experiments are to be repeated many times to have a good comparison of the expected and observed behaviour.

---

<sup>1</sup>There is dispute on whether the idea of the atom started in Greek or Indian philosophy. The Indian philosophers Jain, Ajivika and Carvaka in the 5<sup>th</sup> century BC might have started an epistemological discussion on this subject independently [2].

A second issue could be raised on the value of the expression “widely accepted” for a scientific model. If an assessment of a scientific model is to be objective and within the framework of empiricism, whether it is accepted by a community or not should not affect the critique of any model under study. In this document, this issue is not raised either, and we limit ourselves to the study of the models based on a rational and objective analysis over experimental evidence.

A current model of matter that is able to predict a large amount of phenomena with excellent accuracy is called the Standard Model [4]. It includes a myriad of fundamental elements with a complex interaction between them. The “top” quark is one of the particles in the Standard Model and it interacts through all kinds of forces in the model: the strong interaction, the weak interaction and the electromagnetic interaction. It was discovered at Fermilab, only in 1995 [5,6], with interesting properties, including a large rest mass [7], compared to the other particles in the Standard Model. It belongs to the classification of a “quark” in the Standard Model, of which there are six flavours.

An interesting effect of the fact that quarks interact through the strong force is that, in most cases, quarks cause showers of particles to be produced through a mechanism dominated by the strong interaction. Due to the top quark’s short lifetime, it decays very fast through weak interactions, instead of generating a shower of particles through the strong force, as do other quarks. A study of the strong force radiation emitted in the production and decay of the top quark allows one to clarify a bit more the connection between the top quark and the strong force. Another observed characteristic of the top quark is that it decays very often into the second heaviest quark, the  $b$ -quark [7], which needs to be well detected if one wants to study the top quark.

The Higgs boson [8,9], observed in 2012, plays a central role in the Standard Model [10], particularly in the mechanism of electroweak symmetry breaking (see Chapter 2 for more details). Furthermore, it also couples strongly to the top quark, which proposes that the study of this connection can be a useful way of probing the characteristics of both of these particles. Exploring the properties of both particles is also a helpful guide towards testing other models besides the Standard Model, which predict alternative mechanisms for electroweak symmetry breaking.

Although widely accepted, the Standard Model is not the only theory of matter in particle physics and a large amount of competing theories see the special properties of the top quark as an excellent scenario to extend the Stan-

dard Model’s predictions with fresh ideas of what could happen in unprobed environments. These competing models, frequently referred as being “Beyond the Standard Model”, often expect that unobserved particles have a connection with the top quark.

This thesis focuses on studies on the top quark, taking advantage of its interesting position in the Standard Model to explore its relation to the strong interaction and novel mechanisms by which it could be produced, in the context of models Beyond the Standard Model. The former is done by measuring the production cross section of top-antitop pairs from proton-proton collisions as a function of the number of jets produced by strong force radiation. The latter is achieved in two separate analyses, comparing the invariant mass of the top-antitop system, produced in proton-proton collisions, with the one predicted by the Standard Model or by proposals Beyond the Standard Model.

This document is divided into three main parts. The first part of this thesis is composed only of Chapter 2, which discusses the current understanding of the Standard Model in a brief overview, focusing on its relation with the relevant aspects of the top quark used in the studies in this thesis. The second part focuses on the experimental setup used to perform the analyses. The measurements and searches are done using the results of proton-proton collisions in the ATLAS [11] detector, at the Large Hadron Collider (LHC) [12], which also deserves an introduction in Chapter 3. Chapter 4 shows a few performance studies in the selection of  $b$ -quark-enriched events in ATLAS.

The third part details the three physics analyses performed. Chapter 5 explains the measurement of the top-antitop production cross section as a function of the jet multiplicity in the final state, using data of proton-proton collisions in ATLAS at a center-of-mass energy of 7 TeV. The observed data is corrected for the detector effects in an “unfolding” procedure and a comparison of different simulations of the Standard Model prediction is shown. Chapter 6 discusses the selection of top-antitop pairs produced in proton-proton collisions at center-of-mass energy of 7 TeV, with a focus in probing for Beyond the Standard Model physics. A comparison is done between the Standard Model predicted and the observed spectra for the top-antitop invariant mass. Comparisons are also done between data and alternate models for top-antitop pair production. Chapter 7 extends the previous chapter, by performing a very similar analysis using data from proton-proton collisions at a center-of-mass energy of 8 TeV. Chapter 8 summarises the targets proposed and the results obtained. The appendix contains a few data to simulation comparison

distributions related to the analysis discussed in Chapter 5.

As a final comment, the units used in this document are such that  $\hbar = c = 1$ , so that, in this system, the units of length and time are the same and they are the inverse of the units of energy and mass:

$$[\text{length}] = [\text{time}] = [\text{energy}]^{-1} = [\text{mass}]^{-1} = [\text{momentum}]^{-1}. \quad (1.1)$$

In this system, a particle's mass is numerically equal to its energy in its rest frame  $mc^2$  and its inverse Compton wavelength  $mc/\hbar$ . The Einstein summation convention of summing repeated indices in vectors and tensors is adopted.

# **Part I**

## **Theoretical foundations**

# Chapter 2

## Theory overview

This chapter briefly overviews a few theoretical concepts relevant to understand the goals, methods and results of the physics analyses that follow. This text assumes that the Standard Model of particle physics, discussed in the next sections, is valid up to a good approximation, working as a reasonable effective theory. The methods used in the physics analyses heavily rely on Monte Carlo simulation, which is also reviewed in Section 2.4.

### 2.1 The Standard Model

The best validated model of particle physics so far is the Standard Model (SM) [10, 13–26]. It includes a set of fundamental particles which interact through three forces: the strong nuclear force, the weak nuclear force and the electromagnetic force. The description of the particles and their interactions is done in the framework of a Quantum Field Theory [4]. A set of fields exist in the Standard Model, which model particles with different attributes. Fermions are particles which have a half-integer spin and include “quarks”, fundamental building blocks of the protons and neutrons that are subject to the interaction of the strong force, and “leptons”, which do not interact through the strong force. Bosons are integer spin particles.

The quarks in the Standard Model are the “up”, “down”, “charm”, “strange”, “top” and “bottom” fermions. Furthermore, the three lepton generations include the “electron”, “muon” and “tau”, with their respective neutrinos. The force mediation in the SM is described by the requirement of gauge symmetries, that is, transformations on the fields that do not change the Lagrangian that describes the theory. The gauge symmetry requirement leads to the existence of a set of “gauge boson” fields, which transmit the interaction. The Standard



Model is based on the gauge symmetries  $SU(3)_C \times SU(2)_L \times SU(1)_Y$ . The  $SU(3)_C$  gauge symmetry generates the strong interactions through the “gluon” fields, while the  $SU(2)_L \times SU(1)_Y$  symmetry generates the electroweak interactions, which are transmitted by the  $W^\pm$ , photon and  $Z$  fields. The Higgs field couples to the SM matter particles, the  $W^\pm$  and the  $Z$  bosons, providing them with mass. A summary of some properties of the particles in the SM is shown in Table 2.1. Their masses were rounded and the errors omitted, with the purpose of showing only their order of magnitude in comparison to each other. Note that the top quark has the largest mass among all particles.

Table 2.1: Properties of the fundamental particles of the Standard Model. Information extracted from [7]. Particle’s masses were rounded to show their order of magnitude. Latest measurements including errors can be found in [7].

Quarks	<div>2.4 MeV/c<sup>2</sup></div> <div>2/3</div> <div>1/2</div> <div>u</div> <div>up</div>	<div>1.27 GeV/c<sup>2</sup></div> <div>2/3</div> <div>1/2</div> <div>c</div> <div>charm</div>	<div>172.5 GeV/c<sup>2</sup></div> <div>2/3</div> <div>1/2</div> <div>t</div> <div>top</div>	<div>0 MeV/c<sup>2</sup></div> <div>0</div> <div>1</div> <div>γ</div> <div>photon</div>	<div>~ 125 GeV/c<sup>2</sup></div> <div>0</div> <div>0</div> <div>H</div> <div>Higgs</div>
	<div>4.8 MeV/c<sup>2</sup></div> <div>-1/3</div> <div>1/2</div> <div>d</div> <div>down</div>	<div>104 MeV/c<sup>2</sup></div> <div>-1/3</div> <div>1/2</div> <div>s</div> <div>strange</div>	<div>4.2 GeV/c<sup>2</sup></div> <div>-1/3</div> <div>1/2</div> <div>b</div> <div>bottom</div>	<div>0 MeV/c<sup>2</sup></div> <div>0</div> <div>1</div> <div>g</div> <div>gluon</div>	Bosons
	Leptons	<div>0.511 MeV/c<sup>2</sup></div> <div>-1</div> <div>1/2</div> <div>e</div> <div>electron</div>	<div>105.7 MeV/c<sup>2</sup></div> <div>-1</div> <div>1/2</div> <div>μ</div> <div>muon</div>	<div>1.77 GeV/c<sup>2</sup></div> <div>-1</div> <div>1/2</div> <div>τ</div> <div>tau</div>	<div>91.2 GeV/c<sup>2</sup></div> <div>0</div> <div>1</div> <div>Z</div> <div>Z</div>
<div>unknown but &gt; 0 eV/c<sup>2</sup></div> <div>0</div> <div>1/2</div> <div>ν<sub>e</sub></div> <div>electron neutrino</div>		<div>unknown but &gt; 0 eV/c<sup>2</sup></div> <div>0</div> <div>1/2</div> <div>ν<sub>μ</sub></div> <div>muon neutrino</div>	<div>unknown but &gt; 0 eV/c<sup>2</sup></div> <div>0</div> <div>1/2</div> <div>ν<sub>τ</sub></div> <div>tau neutrino</div>	<div>80.4 GeV/c<sup>2</sup></div> <div>±1</div> <div>1</div> <div>W</div> <div>W</div>	

The electroweak interactions couple differently to right-handed and left-handed fields, which are grouped differently in  $SU(2)$  singlets and doublets representations [4], as follows:

$$\begin{array}{cccccc}
\begin{pmatrix} u \\ d \end{pmatrix}_L & \begin{pmatrix} c \\ s \end{pmatrix}_L & \begin{pmatrix} t \\ b \end{pmatrix}_L & \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L & \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L & \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \\
u_R & c_R & t_R & & & \\
d_R & s_R & b_R & e_R & \mu_R & \tau_R
\end{array}$$

in which the left-handed fields have the  $L$  subscript and they are arranged in the doublet representation of the  $SU(2)$  group, while the right-handed particles are arranged in the singlet representation with the  $R$  subscript. The fields have quantum numbers associated to their interaction in the SM, which will be detailed in the next subsection. For completeness, the fields' quantum numbers for their electroweak interactions is described here. The fields arranged in  $SU(2)$  doublets have their third component of the isospin  $I_3 = \pm\frac{1}{2}$ , with the positive value associated with the up-type quarks and neutrinos and the negative value, to down-type quarks, electron, muon and tauon. The right-handed fields have  $I_3 = 0$ . The up-type quarks have an electric charge  $Q = +\frac{2}{3}$  and the down-type quarks have  $Q = -\frac{1}{3}$ . Neutrinos have neutral electric charge and electron, muons and tau have charge  $Q = -1$ . A third quantum number relates to the weak hypercharge  $Y$  and it is associated such that the relation  $Q = I_3 + Y$  is respected<sup>1</sup>.

Note that there is no mention of the right-handed neutrinos in this discussion, which is how the Standard Model was initially presented, since the neutrinos were assumed to be massless. Experimental evidence suggests that neutrinos are not massless, though, and the theory should be changed in that respect [27–29]. This is not a central point in the analyses that follow, so it will not be further discussed in this document.

The SM is described by a Lagrangian density that can be separated into a gauge term, a matter term, a Yukawa term and a Higgs term [30]:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Matter}} + \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}}.$$

The matter Lagrangian contains the kinetic energy of the quarks and leptons and their interactions, which is given by their coupling to the gauge fields. The gauge term of the Lagrangian includes the kinetic energy of gauge fields for strong and electroweak interactions, leading to the description of their

---

<sup>1</sup>Some authors use the  $Q = I_3 + Y/2$  convention instead. This document follows the convention in [4].

propagation mechanism. The interaction of the Higgs field with the quarks and leptons is done by the Yukawa interaction terms of the Lagrangian, which provides a dynamical mechanism by which the particles acquire mass. Finally, the Higgs field sector contains the Higgs kinetic energy and the Higgs potential, which causes a non-zero vacuum expectation value for the Higgs field.

### 2.1.1 Matter fields and electroweak interactions

The matter fields are generically represented as spinors  $\Psi$ , which could be incorporated as a free field with a single kinetic term [4, 31]:

$$\mathcal{L}_0 = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi, \quad (2.1)$$

in which  $\gamma^\mu$  is defined as a set of four matrices that satisfy  $\{\gamma^\mu, \gamma^\nu\} \equiv \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}I_4$ , where  $g^{\mu\nu}$  is the Minkowski metric [4] and  $I_4$  is the  $4 \times 4$  identity matrix. The Dirac adjoint is defined as  $\bar{\Psi} \equiv \Psi^\dagger\gamma^0$ , where  $\Psi^\dagger$  represents the Hermitian adjoint of  $\Psi$ .

This results in a field that satisfies the Dirac equation [4], but does not include the interactions in the theory. As was mentioned previously, the Standard Model demands gauge invariance over a set of Lie groups, particularly over  $SU(2)_L \times U(1)_Y$  for the electroweak force. This gauge invariance is implemented by demanding the invariance of the Lagrangian on the transformation:

$$\begin{aligned} \Psi_L &\rightarrow e^{i\alpha(x)\cdot\mathbf{T}}\Psi_L, \\ \Psi_R &\rightarrow \Psi_R, \end{aligned} \quad (2.2)$$

for the  $SU(2)_L$  symmetry and:

$$\begin{aligned} \Psi_L &\rightarrow e^{i\beta(x)Y(\Psi_L)}\Psi_L, \\ \Psi_R &\rightarrow e^{i\beta(x)Y(\Psi_R)}\Psi_R, \end{aligned} \quad (2.3)$$

for the  $U(1)_Y$  symmetry. In these equations,  $\Psi_R$  and  $\Psi_L$  represent the right-handed and left-handed fields respectively;  $Y(\Psi)$  is the weak hypercharge of the field  $\Psi$ ;  $\mathbf{T} = \begin{pmatrix} T^1 & T^2 & T^3 \end{pmatrix}^T$  is the weak isospin operator <sup>2</sup>, whose components are the generators of the  $SU(2)_L$  transformation and can be written

---

<sup>2</sup>The symbol  $T$  in the superscript indicates that the transverse of the matrix is to be taken.

as  $T^a = \frac{1}{2}\sigma^a$  (where  $\sigma^a$  are the Pauli matrices);  $\alpha(x)$  is an arbitrary three-component vector of functions of the space-time;  $\beta(x)$  is an arbitrary function of the space-time. The Lagrangian can be made invariant over these transformations, by substituting the  $\partial_\mu$  in  $\mathcal{L}_0$  by the covariant derivative  $D_\mu$ :

$$D_\mu \Psi = (\partial_\mu - ig \mathbf{W}_\mu \cdot \mathbf{T} - ig' Y(\Psi) B_\mu) \Psi, \quad (2.4)$$

in which  $g$  and  $g'$  are coupling constants and gauge fields<sup>3</sup>  $\mathbf{W} = \begin{pmatrix} W^1 & W^2 & W^3 \end{pmatrix}$  and  $B$  are incorporated<sup>4</sup>. It is implicit that only the left-handed fields couple to the  $SU(2)_L$  generators term.

The usage of the covariant derivative  $D_\mu$  includes the coupling of the matter fields with the  $\mathbf{W}$  and  $B$  fields in the kinetic term of the Lagrangian. The gauge fields have their kinetic terms added in the gauge part of the Lagrangian as  $-\frac{1}{4}W_{\mu\nu}^a W^{\mu\nu,a} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu}$ , with:

$$\begin{aligned} W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk} W_\mu^j W_\nu^k, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \quad (2.5)$$

The mass eigenstates of the electroweak fields are not  $W^1$ ,  $W^2$ ,  $W^3$  and  $B$ , but a linear transformation of them, represented as  $W^+$ ,  $W^-$ ,  $Z$  and the photon field  $A$ . The electroweak fields acquire mass through the mechanism mentioned in Section 2.1.3. It is convenient to treat the electroweak fields in their mass eigenstates, given by:

$$W_\mu^+ = \frac{W_\mu^1 - iW_\mu^2}{\sqrt{2}}, \quad (2.6)$$

$$W_\mu^- = \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}}, \quad (2.7)$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad (2.8)$$

where the weak mixing angle  $\theta_W$  is defined such that  $g \sin \theta_W = g' \cos \theta_W = e$ , where  $e$  is the charge of the electron.

---

<sup>3</sup>In this document, the term “gauge fields” is used to refer to these objects. In some (mostly mathematical texts), these objects are called “connections”.

<sup>4</sup>Note that in  $\mathbf{W}_\mu \cdot \mathbf{T}$ , the  $\mathbf{A} \cdot \mathbf{B}$  operation is defined as  $\mathbf{A} \cdot \mathbf{B} \equiv A^a B^a \equiv \sum_a A^a B^a$ .

### 2.1.2 Quantum Chromodynamics

Another force included in the Standard Model is the strong interaction, which is in the domain of Quantum Chromodynamics (QCD). The strong interaction is incorporated by demanding invariance of the Lagrangian due to a  $SU(3)_C$  local gauge transformation:

$$\Psi \rightarrow e^{i\kappa(x)\cdot\mathbf{t}}\Psi, \quad (2.9)$$

in which  $\mathbf{t} = \begin{pmatrix} t^1 & t^2 & t^3 & t^4 & t^5 & t^6 & t^7 & t^8 \end{pmatrix}^T$  are the  $SU(3)$  generators and  $\kappa(x)$  is an arbitrary eight-component vector of functions of the space-time. The gauge invariance requirement is incorporated in the Standard Model, by including the gluon field in the covariant derivative described previously (Equation 2.4):

$$\begin{aligned} D_\mu \Psi &= (\partial_\mu - ig\mathbf{W}_\mu \cdot \mathbf{T} - ig'Y(\Psi)B_\mu)\Psi \\ &\quad - (ig_S\mathbf{G}_\mu \cdot \mathbf{t})\Psi, \end{aligned} \quad (2.10)$$

in which the  $\mathbf{G}$  is an eight-component vector that represents the gluon fields and  $g_S$  is the strong force coupling constant. It is also common to write the strong coupling constant as a function of  $\alpha_S = g_S^2/4\pi$ . This addition to the covariant derivative (for quarks) allows the quark fields to emit and absorb gluons, through the  $i\bar{\Psi}\gamma^\mu D_\mu \Psi$  term of the Lagrangian. The dynamics of the new gluon fields can be included by an extra kinetic term  $-\frac{1}{4}F_{\mu\nu}^a F^{\mu\nu,a}$ , where:

$$F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_S f^{abc} G_\mu^b G_\nu^c. \quad (2.11)$$

Calculations in a Quantum Field Theory can be done by ignoring the interaction terms for the fields at first, and including them at a later stage as a perturbation to the free field solutions, taking only the first terms in the expansion [4]. When calculating the terms of the perturbative expansion in both QCD and the electroweak theory, there are ultraviolet [4] divergencies<sup>5</sup>, which lead to a non-physical prediction for some observables. In “renormalisable” theories [4], such as QCD and the electroweak theory, these divergencies happen only in a countable number of interaction diagrams and they can be removed from the physical observables by redefining the Lagrangian so that a

---

<sup>5</sup>We use the jargon “ultraviolet divergencies” here to refer to divergencies caused by high energies and “infrared divergencies” to refer to divergencies caused by low energies.

new set of parameters have these divergencies “subtracted”. The process of redefining the model in such a way that these divergencies are removed from physical observables is called “renormalisation” [4]. It can be done in multiple ways, leading to different “renormalisation schemes” [4]. The renormalised parameters of the theory depend on a “renormalisation scale”  $\mu_R$ , although the physical observables do not depend on this scale.

A consequence of this procedure in QCD is that the renormalised coupling constant  $\alpha_S$  depends on the scale  $\mu_R$  and on the (squared) momentum transfer in a QCD interaction  $-Q^2$ . For low energies, QCD has a non-perturbative behaviour and the perturbative expansion cannot be made. In such situations, other tools exist to study the QCD theory, such as Lattice QCD [32]. For higher energies though, the renormalised QCD coupling constant  $\alpha_S(\mu_R)$  is such that the perturbative expansion is valid, as it can be seen in its leading-order approximation:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu_R^2)}{1 + (b_0\alpha_S(\mu_R^2)/2\pi) \ln(Q^2/\mu_R^2)}, \quad (2.12)$$

where  $b_0$  is a positive coefficient [4] (assuming the approximation of massless quarks) and, therefore, for high enough values of  $Q^2$  QCD can be analysed using perturbation theory, in what is often called “pQCD” (for “perturbative QCD”). pQCD can be used for interactions involving momentum transfers well above the  $\Lambda_{\text{QCD}} \sim 200$  MeV experimental threshold, for example, above  $Q = 1$  GeV, where  $\alpha_S(Q^2) \sim 0.4$  [4].

With this behaviour (often called “asymptotic freedom”), the high energy environment is an ideal scenario to study QCD, since perturbation theory provides an excellent tool to study its interactions. One of the key topics of this thesis is to measure the effect of QCD radiation at high energies, produced in association with the top quark.

For smaller  $Q^2$  though, the coupling constant increases and QCD grows even stronger. In this scenario, the strong force also becomes stronger as the distance between the particles increase. At sufficiently large distances, the potential energy between the quarks, produced by QCD, is strong enough to produce new quark pairs. This effect, called “confinement” does not allow one to measure a free quark: they are always in colourless bound states. What can be measured in the detector, therefore, is never a quark itself, but its byproducts. For this reason, the QCD radiation measured in this document is detected as a set of particles in a region of the experiment, and not as bare quarks or gluons.

### 2.1.3 Electroweak symmetry breaking mechanism

The sectors of the Standard Model discussed previously include fields for the matter particles and the gauge boson fields  $W^\pm$ ,  $Z$ , the photon and the gluons. While the photon and the gluons are massless, the electron, muon, tauon, the six quarks, the  $W^\pm$  and  $Z$  are not <sup>6</sup>. Including mass terms in the Lagrangian which resemble  $m\bar{\Psi}\Psi$  or  $mV_\mu V^\mu$  can be shown to violate the gauge invariance of the SM [4].

A mechanism has been devised, called the “Englert - Brout - Higgs - Guralnik - Hagen - Kibble mechanism” <sup>7</sup>, which introduces a scalar field [4]  $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ , which is an  $SU(2)$  doublet. This field couples to the gauge bosons described previously due to its kinetic term and it couples to itself through the potential  $V(\phi)$ :

$$\mathcal{L}_{\text{Higgs}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi) \text{ , where:} \quad (2.13)$$

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \frac{\lambda}{2} (\phi^\dagger \phi)^2, \quad (2.14)$$

in which  $\mu^2$  and  $\lambda$  are parameters of the model. For  $\mu^2 > 0$ , the potential has a maximum at the origin and minima at  $\langle \phi \rangle_0 = v = \left(\frac{\mu^2}{\lambda}\right)^{1/2}$ . While the point at zero is unstable, the minima are stable and sets the vacuum expectation value for the field.

Gauge invariance is preserved in the Lagrangian, through this mechanism, and the coupling of this field to the fermion and gauge boson fields dynamically generates their mass. The interaction terms between the Higgs boson and the matter fields are added in the Yukawa sector of the Lagrangian. Note as well that the gauge bosons have their masses included by the  $|D^\mu \phi|^2$  term of the Lagrangian.

A further effect of this spontaneous symmetry breaking of the Lagrangian is the existence of a Higgs boson particle with mass  $\sqrt{2}\mu$ . A Higgs boson has recently been verified experimentally at the LHC [8, 9] and it is still under

---

<sup>6</sup>As it was already mentioned, there is evidence that supports that neutrinos are not massless, although their mass is much smaller than any other particle in the Standard Model. They are assumed massless here though, since their mass is much smaller than what can be experimentally verified in the ATLAS detector (Chapter 3) and their non-zero mass does not affect the process of the measurement and searches shown in this document.

<sup>7</sup>Other names are also used for it, including the “Higgs mechanism”, the “Brout – Englert – Higgs mechanism” and the “ABEGHHK’tH mechanism” (for Anderson, Brout, Englert, Guralnik, Hagen, Higgs, Kibble and ’t Hooft).

study. Note that the masses associated to the matter fields are related to their coupling to the Higgs field, therefore, the most massive particle in the Standard Model, the top quark, would be strongly connected to the Higgs boson.

## 2.2 The Standard Model and the top quark

The top quark, discovered in 1995 at Fermilab, has a few interesting properties which are going to be explored in this document. To discuss the top quark, a few remarks will be made first on a few elements of the Standard Model interactions.

The electroweak interaction can be expressed in terms of the  $W^\pm$ ,  $Z$  and photon fields, but only the  $W^\pm$  currents change fermions' flavours at tree level, while the neutral bosons do not change the fermion flavour. Decays of particles which change flavour and include neutral currents are allowed in the SM in diagrams with loops, but these are highly suppressed [15]. One important characteristic of the SM is that the  $W^\pm$  bosons do not act on the quark fields described previously, but on a linear combination of quark fields, with weights given by the CKM matrix [25,26,33]. It is instructive to verify how the charged currents  $J_W^\mu$  (and its adjoint  $J_W^{\mu\dagger}$ ) and the neutral currents  $J_Z^\mu$  and  $J_{EM}^\mu$  show the interaction between the fermions and the  $W^\pm$ -bosons after the electroweak symmetry breaking [4,33]:

$$\mathcal{L}_{\text{SM}} \supset g \left( W_\mu^- J_W^\mu + W_\mu^+ J_W^{\mu\dagger} + Z_\mu J_Z^\mu \right) + e A_\mu J_{\text{EM}}^\mu, \quad (2.15)$$

where the symbol  $\supset$  indicates that the Lagrangian can be rewritten so that this is one of its terms [4]. The currents can be defined as follows [4]:

$$J_W^{\mu\dagger} = \frac{1}{\sqrt{2}} (\bar{\nu}_L \gamma^\mu \ell_L + \bar{U}_L \gamma^\mu V_{CKM} D_L), \quad (2.16)$$

$$\begin{aligned} J_Z^\mu = \frac{1}{\cos \theta_W} & \left[ \bar{\nu}_L \gamma^\mu \left( \frac{1}{2} \right) \nu_L + \bar{\ell}_L \gamma^\mu \left( -\frac{1}{2} + \sin^2 \theta_W \right) \ell_L + \bar{\ell}_R \gamma^\mu (\sin^2 \theta_W) \ell_R + \right. \\ & + \bar{U}_L \gamma^\mu \left( \frac{1}{2} - \frac{2}{3} \sin^2 \theta_W \right) U_L + \bar{U}_R \gamma^\mu \left( -\frac{2}{3} \sin^2 \theta_W \right) U_R + \\ & \left. + \bar{D}_L \gamma^\mu \left( -\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W \right) D_L + \bar{D}_R \gamma^\mu \left( \frac{1}{3} \sin^2 \theta_W \right) D_R \right], \end{aligned} \quad (2.17)$$

$$J_{\text{EM}}^\mu = \bar{\ell} \gamma^\mu (-1) \ell + \bar{U} \gamma^\mu \left( +\frac{2}{3} \right) U + \bar{D} \gamma^\mu \left( -\frac{1}{3} \right) D \quad (2.18)$$

$$(2.19)$$



in which the matter fields are arranged for convenience such that (the same symbols are used to identify the particles as in Table 2.1) <sup>8</sup>:

$$\nu_L = \begin{pmatrix} \nu_e & \nu_\mu & \nu_\tau \end{pmatrix}_L^T, \quad (2.20)$$

$$\ell_L = \begin{pmatrix} e & \mu & \tau \end{pmatrix}_L^T, \quad (2.21)$$

$$\ell_R = \begin{pmatrix} e & \mu & \tau \end{pmatrix}_R^T, \quad (2.22)$$

$$U_L = \begin{pmatrix} u & c & t \end{pmatrix}_L^T, \quad (2.23)$$

$$D_L = \begin{pmatrix} d & s & b \end{pmatrix}_L^T, \quad (2.24)$$

$$U_R = \begin{pmatrix} u & c & t \end{pmatrix}_R^T, \quad (2.25)$$

$$D_R = \begin{pmatrix} d & s & b \end{pmatrix}_R^T, \quad (2.26)$$

$$\ell = \ell_L + \ell_R, \quad (2.27)$$

$$U = U_L + U_R, \quad (2.28)$$

$$D = D_L + D_R, \quad (2.29)$$

and the  $V_{CKM}$  matrix is the CKM [25, 26] matrix given by:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (2.30)$$

An important side effect of this structure for the Lagrangian density is that the coupling of the charged currents to the quarks includes elements of the  $V_{CKM}$  matrix, which have been experimentally measured [7]. A key characteristic of the CKM matrix [7] is that it has a  $|V_{tb}|$  element very close to one, while the elements  $|V_{td}|$  and  $|V_{ts}|$  are very close to zero. The Feynman diagram vertex for the interaction between the top quark and the  $W$ -boson involves a  $|V_{tb}|$  contribution which enhances the decay of the top quark into a  $W$ -boson and a  $b$ -quark, while the small values of the other elements in the same row suppress the decay of the top quark into the other flavours. In fact, the branching ratio of top decays into a  $W$ -boson and a  $b$ -quark has been measured to be  $0.99^{+0.09}_{-0.08}$  (see [7]).

Furthermore, the neutral currents that describe the interaction of the  $Z$ -boson and the photon field with the quarks do not change the flavour of the

---

<sup>8</sup>The symbol T in the superscript represents the transpose of the matrices.

fermions [15], so that at tree-level, the top quark cannot decay into another quark interacting through a  $Z$ -boson or a photon. It can decay using neutral currents through diagrams including loops, but these are highly suppressed, compared to the charged current tree-level diagram. There are analyses that search for higher rates for a single-top production through Flavour Changing Neutral Currents (FCNC) and the current limits are given by the ATLAS search [34], which has shown good agreement between data and the SM. The current cross section limit on the FCNC single top production is  $\sigma_{qg \rightarrow t} \times \text{Br}(t \rightarrow Wb) < 3.9 \text{ pb}$ .

Also, the top quark has a lifetime of  $\sim 5 \times 10^{-25} \text{ s}$  [30, 35], which is smaller than the characteristic formation time of hadrons  $\tau_{\text{form}} \sim 1/\Lambda_{QCD} \sim 3 \times 10^{-24} \text{ s}$  [30], where  $\Lambda_{QCD}$  is the scale at which QCD becomes non-perturbative. This means that the top quark does not hadronise and that its decay happens almost exclusively through the  $t \rightarrow Wb$  channel at tree-level, with other contributions being suppressed.

A study of the top quark radiation allows for a deeper understanding of the Standard Model, particularly the QCD interaction. It also helps understanding and verifying the Standard Model description of this quark. More information can be found in a review of the Standard Model with a focus in the top quark and its discovery, in [30].

## 2.3 Top-antitop pair generation at the LHC

The top quark is produced predominantly in the LHC proton-proton collisions through strong interactions, paired with its anti-quark, as it is shown in Figure 2.1. The top pair production in the Tevatron experiments [5, 6, 36], which collide protons and antiprotons, is dominated ( $\sim 85\%$  of the cases) by the quark-antiquark diagram (c) in Figure 2.1, while in the LHC, the gluon-gluon diagrams (a), (b) and (d) have the dominant contributions ( $\sim 90\%$  of the cases). The result of a full calculation of the top pair production cross section can be seen in the review in [31].

As mentioned previously, due to the large value of the  $|V_{tb}|$  matrix element, each top quark decays almost exclusively to  $t \rightarrow Wb$ . The  $W$  boson may then decay leptonically, into a lepton and a neutrino, or hadronically, into quarks, producing jets in the detector. This is shown, schematically, in Figure 2.2.

These decays create three different channels of study of the  $t\bar{t}$  system:

- $t\bar{t} \rightarrow \ell\bar{\ell}\nu\nu b\bar{b}$ : the dilepton channel, which assumes that both  $W$  bosons

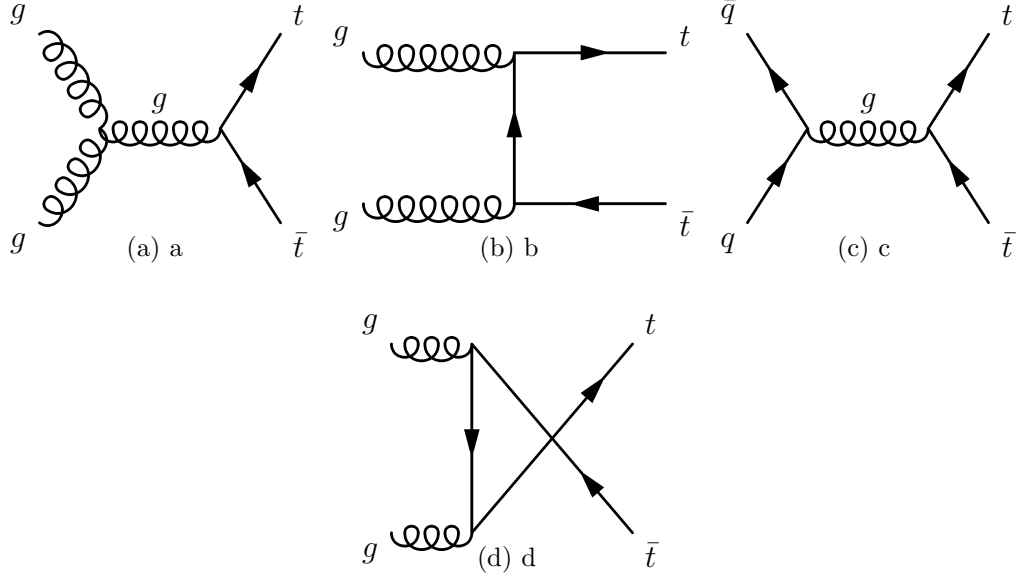


Figure 2.1: Top-antitop main production diagrams at the LHC, at tree-level.

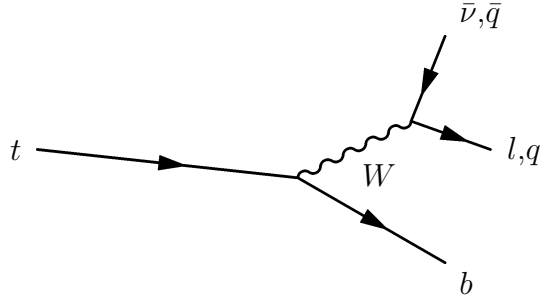


Figure 2.2: Top quark decays

decay to leptons and neutrinos;

- $t\bar{t} \rightarrow \ell\nu b\bar{b}jj$ : the single lepton channel, which assumes that one  $W$  boson decays to lepton and a neutrino, while the other one decays hadronically;
- $t\bar{t} \rightarrow b\bar{b}jjjj$ : the all-hadronic channel, which assumes that both  $W$  bosons decay hadronically.

This thesis concentrates on the single lepton (or “semileptonic”) channel.

The top quark may also be produced separately from an antitop, by the production diagrams shown in Figure 2.3. These productions have lower cross sections than the  $t\bar{t}$  system production in the LHC. Note that all diagrams require a  $W$  boson interaction, that is, the electroweak interaction must be present, while top pairs are produced through the strong interactions.

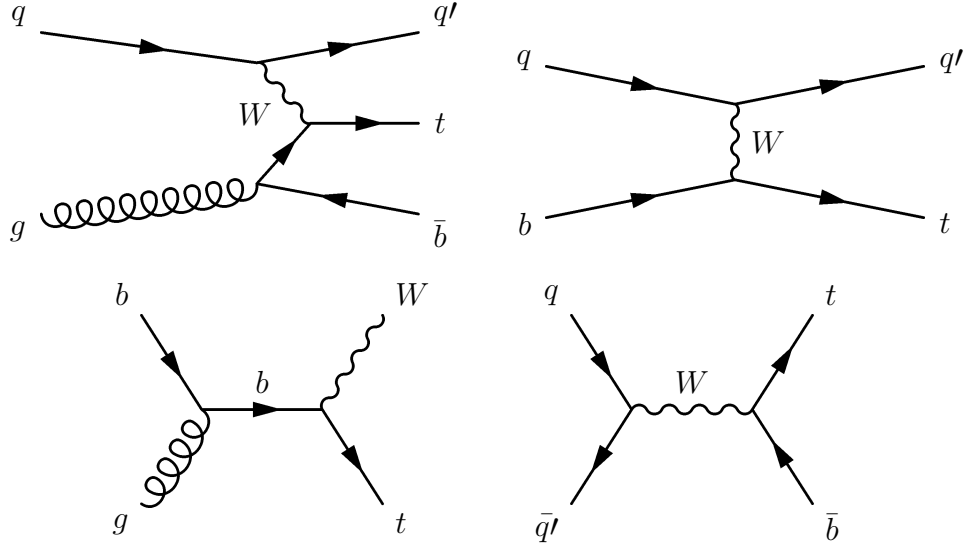


Figure 2.3: Single top production diagrams at tree-level.

## 2.4 Monte Carlo event generators

An important step for the analysis of the LHC data is the simulation of the results, assuming that a certain model is correct. This description is paramount to estimate the behaviour of the well known physics, so that their effect in the measurements or searches can be accounted for. It is equally important to simulate alternative physics models, so that a comparison of the measurements and searches in data and simulation can be used to test proposed extensions of the Standard Model. A comparison of the data and the simulated result might show that the model is compatible with experimental evidence, that a fine-tuning of some of the model's constants is necessary, or that the model is completely inconsistent with empirical data. As a result, in this section a brief description of the methods used for simulation is given. A more complete review can be read in [37]. The simulation uses Monte Carlo methods, by which a probability density function (PDF) is calculated and pseudo-random samples are generated according to that PDF. The prediction of the PDF is given by the model under study using a few assumptions from validated parts of the Standard Model.

A set of steps are required to simulate different parts of the Standard Model, which are described in the next sections. They are summarised graphically in Figure 2.4. Section 2.4.1 describes the Factorisation Theorem and how it fits with the parton-level calculation to provide the Hadron-level simulation result. The Parton Shower simulation is described in Section 2.4.2. The previous steps can be extended with next-to-leading-order calculations, which are mentioned

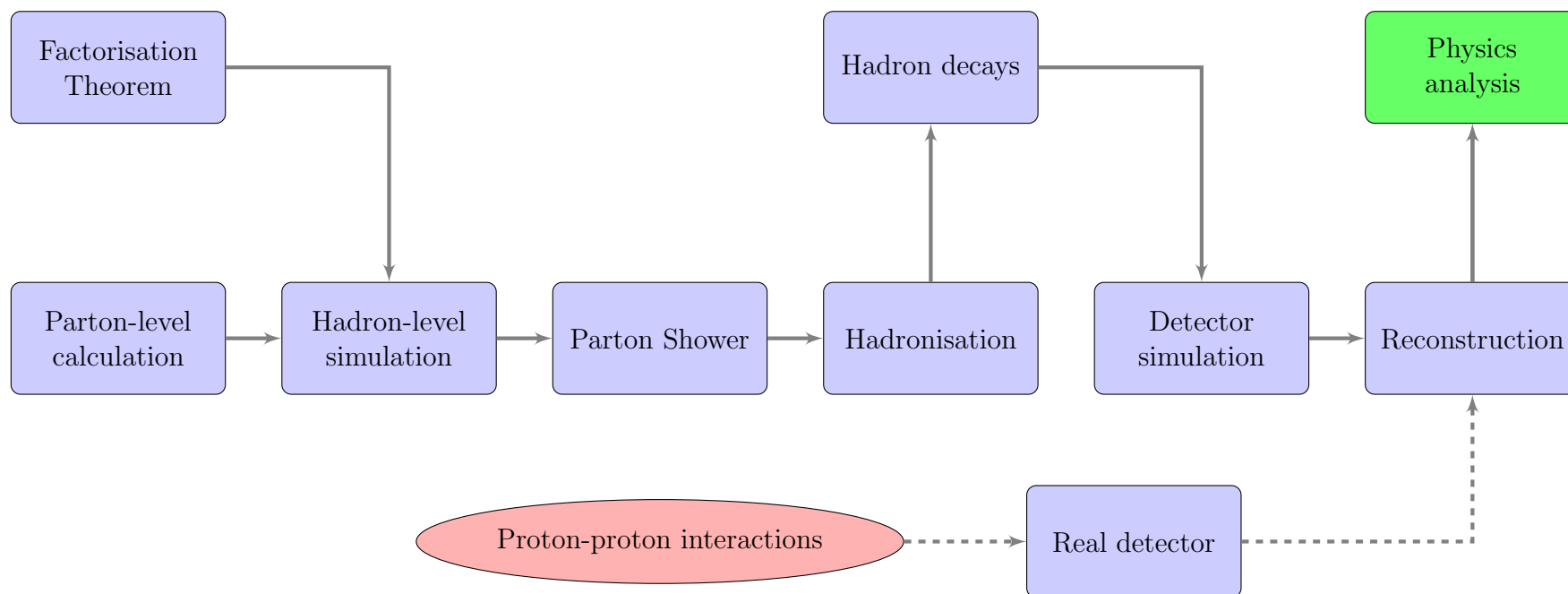


Figure 2.4: Simplified schematic view of simulation steps necessary for physics analyses.

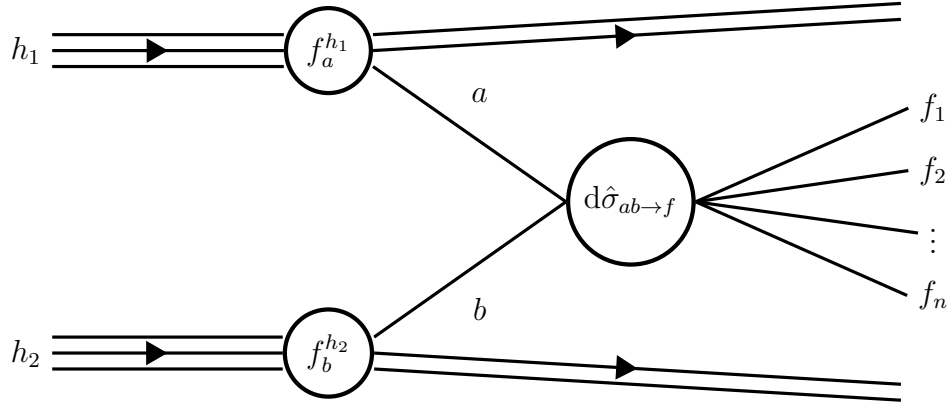


Figure 2.5: Graphical representation of the Factorisation Theorem, Equation 2.31. The symbols are the same as in Equation 2.31, except that  $f_1, f_2, \dots, f_n$  represent multiple particles in the final state  $f$ .

in Section 2.4.3. The Hadronisation and Hadron decay steps are discussed in Section 2.4.4.

### 2.4.1 Factorisation theorem and perturbative treatment

The description of the scattering starts from the Factorisation Theorem, which maps the parton-level cross sections for the quarks and gluon interactions (in the top quark production it follows the leading-order diagrams in the previous section and higher level corrections) into a cross section for the proton-proton collision [37]:

$$\sigma = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b \int f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow f}(x_a p_1, x_b p_2, Q^2, \mu_F, \mu_R), \quad (2.31)$$

where  $f_{a,b}^{h_1,h_2}(x_{a,b}, \mu_F)$  are the parton distribution functions (PDFs<sup>9</sup>), which depend on the momentum fraction  $x_{a,b}$  of the parton  $a, b$  with respect to the protons  $h_1, h_2$ , and on the factorisation scale  $\mu_F$ ;  $\hat{\sigma}_{ab \rightarrow f}$  is the parton-level cross section for the production of the final state  $f$  through the partons  $a, b$ ;  $p_{1,2}$  are the momenta of the protons,  $\mu_R$  is the renormalisation scale used [4]. A graphical representation of Equation 2.31 is shown in Figure 2.5.

The physics result does not depend on the choice of the factorisation scale  $\mu_F$ , which sets the limit between hard processes and soft non-perturbative

---

<sup>9</sup>Although the acronym for PDF can be used to mean “parton distribution function” or “probability density function”, usually this is clear by the context. When it is not, it will be mentioned explicitly.

QCD, or on the choice of the renormalisation scale  $\mu_R$ . The parton distribution functions model the non-perturbative behaviour of the partons in the proton. PDFs for the proton are fitted using data by different groups, such as CTEQ [38, 39], NNPDF [40] and MSTW [41].

The parton-level cross section is calculated for the process generation from the matrix element of the theory using perturbation theory. The equation above can be used in Monte Carlo generator programs, which use the resulting probability density function and pseudo-random number generators to generate events with a certain final-state.

The  $d\hat{\sigma}_{ab \rightarrow f}$  function must be evaluated based on the matrix element (squared) for the scattering and the phase space element over the final state. The matrix element can be calculated from the Feynman diagrams in perturbation theory [4]. An expansion in the QCD sector would resemble:

$$d\hat{\sigma}_{ab \rightarrow f} = C_{LO}\alpha_S^n + C_{NLO}\alpha_S^{n+1} + C_{NNLO}\alpha_S^{n+2} + \dots, \quad (2.32)$$

where  $\alpha_S$  is the strong coupling constant and the coefficients  $C_{LO}$ ,  $C_{NLO}$  and  $C_{NNLO}$  are the leading-order, next-to-leading-order and next-to-next-to-leading-order terms of the perturbative expansion. This expansion can be calculated similarly to include higher order corrections of the electroweak sector. The calculation and Feynman diagrams can then be categorised as being leading-order, next-to-leading-order and henceforth, depending on the coupling constant coefficient in the series. Fixed order Monte Carlo generators calculate the expansion in Equation 2.32 up to a certain maximum exponent on the coupling constant and tree-level expansions do not consider Feynman diagrams involving loops.

This approach is followed in matrix element generators, such as MadGraph [42] at leading-order, but the number of Feynman diagrams, even at tree-level, grows very fast with the final-state particle multiplicity. AcerMC [43] uses MadGraph to generate the matrix element from Feynman diagrams, optimising its procedure to obtain faster results. Another approach is taken by AlpGen [44], which generates tree-level matrix elements numerically with a recursive algorithm, called Alpha [45]. AlpGen's recursive approach allows it to generate multiple particles in the final state.

These methods suffer because the tree-level diagrams diverge whenever external partons become soft or collinear. Similar divergences appear in loop diagrams. For sufficiently inclusive measurements both set of divergences cancel, but for exclusive quantities some logarithmic terms remain, with the form

$L = \ln(Q^2/Q_0^2)$ , where  $Q_0$  is a constant (see Section 2.4.2 for more details). An alternative and complementary approach to the expansion above is to arrange the series in a format that depends on  $\alpha_S L$ :

$$d\hat{\sigma}_{ab \rightarrow f} = \sum_n a_n (\alpha_S L)^n + \alpha_S \sum_n b_n (\alpha_S L)^n + \dots \quad (2.33)$$

The first sum over  $n$  is called the leading logarithm (LL) approximation, while the second sum is the next-to-leading-logarithm one. These terms are incorporated in parton shower simulators, described in the next section, which are able to treat the soft and collinear divergences. The final Monte Carlo simulation is implemented by uniting the power of the parton shower and the fixed-order or tree-level generators.

### 2.4.2 Parton showers

The matrix element generators described previously are quite good in describing high energy interactions, however, as it was mentioned previously, the final state particles also generate soft and collinear branchings, which are best described with parton shower simulations [37]. The parton shower simulation includes the effect of collinear radiation, by calculating approximately the probabilities for parton splitting into more partons at all orders in perturbation theory. The parton shower calculation is valid in the collinear approximation, in which the angle  $\theta$  between the partons after splitting is close to zero. A schematic drawing of a collinear emission is shown in Figure 2.6, in which a parton  $a$  emits an almost collinear parton  $b$  and proceeds as  $b'$ . The variable  $t$  represents an evolution variable, that is, the energy scale in which the splitting occurs, and it could be the four-momentum squared  $p^2$ , for example. The variable  $z$  represents the fraction of energy carried by the parton  $b$ , that is,  $z = E_b/E_a$ , where  $E_a$  is the energy of  $a$  and  $E_b$  is the energy of  $b$ .



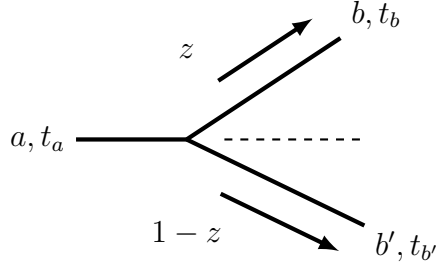


Figure 2.6: Simplified schematic of a parton splitting in which a parton  $a$  emits a parton  $b$  and proceeds as  $b'$ .

Using the DGLAP [46] equations, the probability for this splitting can be written as [47]:

$$\mathcal{P}(a \rightarrow b(z) + b'(1-z) \text{ when } t \rightarrow t + \delta t) = \frac{\delta t}{t} \frac{\alpha_S}{2\pi} P_{bb'}(z), \quad (2.34)$$

where  $P_{ab}$  represents the probability of a parton  $a$  splitting into  $b$ . The idea is that, after the matrix element calculation, a possible splitting (or branching) of an outgoing quark ( $q$ ) or gluon ( $g$ ) can be calculated as [37, 47]:

$$d\hat{\sigma} \sim \sigma_0 \sum_{\text{partons } i} \frac{\alpha_S}{2\pi} \frac{dt}{t} dz P_{ji}(z, \phi) d\phi, \quad (2.35)$$

where  $P_{ji}$  represents the probability of  $j$  splitting into  $i$  (which can be quarks or gluons),  $\sigma_0$  is the hard process matrix element,  $\alpha_S$  is the strong force coupling constant and  $\phi$  is the azimuthal angle in the transverse plane.

In general the splitting functions do not depend on the azimuthal angle, although they need to be included, since they are one of the phase space degrees of freedom. If the splitting functions do not depend on  $\phi$ , this integration amounts to  $2\pi$ . However, if a specific helicity for the initial or final state is to be considered, there could be a dependence on the transverse plane angle.

The probability of the hard process and splitting are factorised in this approximation, so these elements can be considered separately. The parton shower algorithm calculates the probability of not having a branching for a parton with square four-momentum  $q^2$  as  $\Delta_i(q^2, Q_0^2)$ , which can be used in a Monte Carlo algorithm and it is [37, 46]:

$$\Delta_i(q^2, Q_0^2) = \exp \left\{ - \sum_{\text{partons } j} \int_{Q_0^2}^{q^2} \frac{dk^2}{k^2} \frac{\alpha_S}{2\pi} \int_{Q_0^2/k^2}^{1-Q_0^2/k^2} dz P_{ji}(z) \right\}, \quad (2.36)$$

where  $Q_0^2$  is a cut-off for resolvable branchings. The  $\Delta_i(q^2, Q_0^2)$  function is called the Sudakov form factor and it takes advantage of the unitarity relation to include indirectly, all orders in perturbation theory in the collinear approximation.

The parton shower algorithm, uses the  $\Delta_i(q^2, Q_0^2)$  function as follows. A pseudo-random number  $\rho$  is generated with uniform distribution and the equation  $\Delta_i(Q^2, q^2) = \rho$  is solved for  $q^2$ , assuming a maximum virtuality  $Q^2$ . If the solution  $q^2$  is greater than a cut-off for resolvable branchings  $Q_0^2$ , a new branching is created, otherwise, the evolution terminates, since there are no further resolvable splittings. This algorithm implements branchings due to collinear radiation, but implementing soft wide-angle radiation does not seem, at first sight, possible. It can be shown [37], however, that multiple soft wide angle emissions in the parton shower can be implemented by starting with the widest angle emissions and then successively decreasing the angle of the next emissions. This approach, which orders emissions by their angle, is called angular ordered parton shower and it is the one implemented in Herwig [48, 49]. This method for parton showers, with a few modifications, is also used to simulate initial state radiation, before the interaction, since there are also quark and gluon splittings at that stage.

The Pythia [50]<sup>10</sup> implementation of the parton shower is based on a different argument than the angular ordered parton shower. Assuming a large number of colours for an  $SU(N)$  model (in QCD,  $N = 3$ ), it can be observed that partons which have the same colour (they are “colour-connected”) shower independently. The parton shower in Pythia takes advantage of this fact, by creating a “colour flow” connecting the partons and showering the partons in this colour flow independently. The parton shower in this procedure uses transverse momentum ordering, instead of using the angular ordering, as in Herwig [37].

Although the matrix element generators and the parton shower procedures have different strengths, they cannot be combined blindly. Both the matrix element generators and the parton showers include overlapping terms in the perturbative expansion, which would be double counted if a careful method to avoid it is not used. Both Herwig and Pythia include “matrix element corrections” to address this problem consistently. A procedure, called CKKW [51], partitions the phase space into a parton shower dominated region and a matrix

---

<sup>10</sup>Pythia version 8 and the latest releases of Pythia 6 implement this procedure as well as the angular-ordered parton shower mentioned previously.

element dominated region. Events in the matrix element dominated region are “vetoed” in the parton shower description. For these vetoed parton showers, the matrix element is weighted by a Sudakov suppression factor, which includes the probability of non-branching at the same order as in the parton shower. This avoids non-smooth transitions between the parton shower-dominated region and the matrix element-dominated region. An extra weight is used to match the  $\alpha_S$  scales in the parton shower and in the matrix element regions. For the parton shower-dominated region, the showers are also vetoed if they generate splittings that would result in an angular separation bigger than the matrix element partons separation. This avoids extra jets in the final state by construction.

Another method for the matrix element to parton shower matching is the MLM matching [52]. It starts by generating events with the matrix element generator using acceptance cuts and demanding that the hard particles generated are well separated from each other and only allows for branchings consistent with the colour structure of the event. The scale used for  $\alpha_S$  in the matrix element is calculated based on the transverse momentum and on the directions of the momenta for the final state partons. A jet algorithm (see Section 3.5 for definitions of jet algorithms) is applied to the final state partons and, starting from the hardest parton, a geometrical match between the partons and the centroid of the jets is attempted. If a parton does not match any jets, the event is discarded. The events that do pass this veto are required not to have any extra jets, compared to the desired final parton multiplicity used in the matrix element generator: if they do contain extra jets, they are rejected. The last step is equivalent to the Sudakov factor reweighting in the CKKW method. The MLM procedure is used with the Alpgen generator.

Notice that the key difference between the MLM matching scheme and the CKKW one is that, in the CKKW scheme, instead of removing the event if an extra hard jet is available in the event after the parton shower, the matrix element for the extra parton is weighted by a Sudakov suppression factor for the probability of that parton not generating a splitting at the matrix element level (which allows for a smooth transition towards the parton shower). Furthermore, in CKKW the parton shower splittings are vetoed, if they generate new partons with larger angular separation than the separation between the matrix element-level partons. In MLM, the event is showered, without any veto in the parton shower procedure and the full event is vetoed if a parton does not match the final-state jets.

### 2.4.3 Next-to-leading order matrix element generators

Loop diagrams and non-collinear emissions beyond the tree-level diagrams are not included in the tree-level matrix element generators and on the parton shower methods described previously. These diagrams are quite important for top quark physics, to better estimate both the extra radiation effects due to the QCD contribution and to have a better estimate of the inclusive cross section. The next-to-leading-order diagrams include virtual contributions, which add virtual particles to the leading-order diagrams without extra final state particles, and real contributions, which add final state particles to the leading-order diagrams.

For the top-antitop production, a few next-to-leading order corrections are shown in Figure 2.7, which contribute to the  $\alpha_s^3$  expansion of the perturbation series. Graphs (b) and (d) include virtual corrections, while (a) and (c) show real gluon emissions. It is important to note that these corrections interfere with each other and the leading-order diagrams.

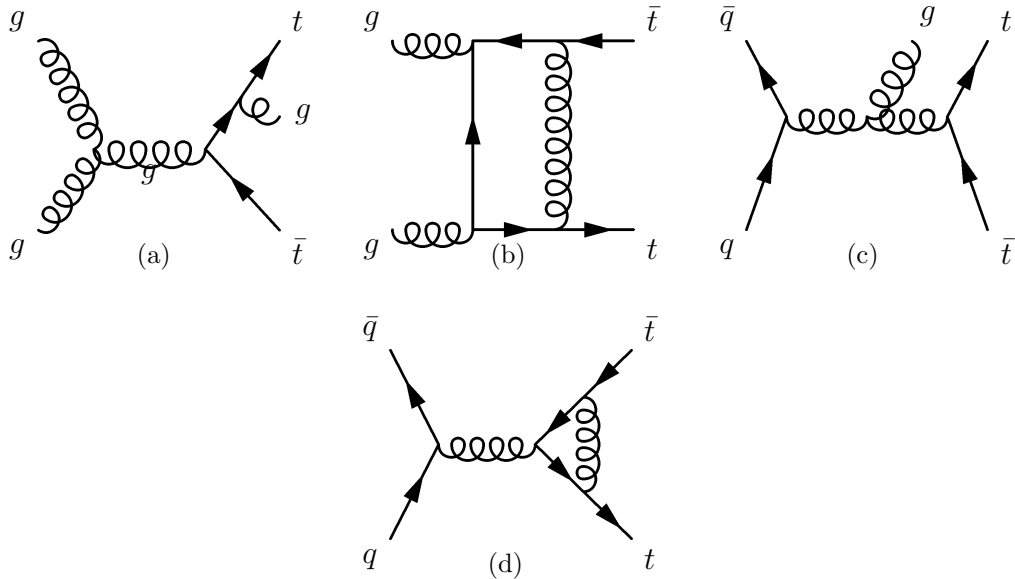


Figure 2.7:  $t\bar{t}$  next-to-leading-order diagram examples.

The matrix element generators described in Section 2.4.1 were limited to the leading order or tree-level generators, while a set of next-to-leading-order (NLO) matrix element generators are currently available for some processes, such as top pair production. Tree-level generators maintain a leading-order precision, but they calculate more correctly differential distributions sensitive to real QCD emission, even at several orders beyond the leading-order calculation, whereas next-to-leading-order calculations are correct in shape and

normalisation at NLO for inclusive variables, but they rely on the parton shower for all extra emissions, beyond the first one [53].

The main problem, for the NLO generators is the matching scheme between the matrix elements and the parton shower, since the latter includes approximate NLO corrections in the Sudakov form factor. Powheg [54] and MC@NLO [55–57] are different NLO generators and they deal with this issue in different ways.

The NLO event contains a leading-order (Born) term corresponding to no emissions, a virtual term due to emissions in loops, and a real term, for final state emissions. These terms can be re-arranged, so that the poles in the virtual and real terms of the NLO calculation are subtracted. The method used in MC@NLO leads to a weighting scheme with negative weights being generated as a result of the way in which this subtraction is implemented. MC@NLO is also only implemented with the parton shower using Herwig, which is another drawback.

The Powheg approach does not include negative weights and it emits the highest transverse momentum parton first, generating the matrix element at NLO. Powheg can also be interfaced with different parton shower generators, as long as the shower is ordered by the hardest emission first. Parton showers with angular ordering can also be implemented, as long as the showering is truncated appropriately [58].

#### 2.4.4 Hadronisation

The parton shower simulation described previously stops when a cut-off scale for the partons' virtuality  $Q_0$  is reached, because at that stage, hadrons are created from the partons and perturbative calculations are no longer valid. This stage is called “hadronisation”<sup>11</sup> and its results cannot be calculated or described using first principles, therefore one must use phenomenological models to simulate it. Two models are mainly used currently: the string model and the cluster model.

In the string model [59], the QCD interaction is modelled as a potential, that attracts partons, creating a colour flux tube between them. After the potential rises above a threshold for a parton creation, a new parton pair is created. An example can be given for a quark-antiquark pair  $q\bar{q}$ , that moves apart, increasing the potential between them. At some point another quark-antiquark

---

<sup>11</sup>The term “hadronisation” can also be used in other contexts. In this document it will refer to the stage in which hadrons are formed, after the parton shower.

pair is created  $q'\bar{q}'$ , reducing the energy of the system, which leads to the formation of the mesons  $q\bar{q}'$  and  $q'\bar{q}$ . This model can also generate baryons, by including the concept of “diquark”, on which, occasionally diquark-antidiquark pairs are produced, instead of quark-antiquark pairs<sup>12</sup>. This model is used in Pythia [50].

The cluster model [60] is based on a property of parton showers called “preconfinement”. The key element in the preconfinement is that the colour structure of the shower at any scale  $Q_0$  is such that any colour singlet combinations of partons (“clusters”) can be formed with an asymptotically universal invariant mass distribution, that is, it depends only on  $Q_0$  and on the QCD scale  $\Lambda$ , but not on the scale  $Q$  of the parton or the hard process initiating the shower, and it has  $Q \gg Q_0$ . If, in addition,  $Q_0 \gg \Lambda$ , the mass distribution of these colour singlet clusters, their momentum and multiplicity can be computed perturbatively [37]. The hadronisation model uses that to enforce gluon splittings into quark-antiquark pairs at the shower cut-off scale  $Q_0$ , so that the adjacent colour lines can be clustered to form mesons. The same concept of “diquark” can be used, as in the string model, to create baryons. It is implemented in Herwig [48, 49].

As a last step, after the hadron formation is simulated, their decays must also be implemented based on observed decay rates, which have been collected by the Particle Data Group [7]. This is the last step of the simulation, before the interaction with the detector, which is done, for the physics analyses in this document, using the software called Geant 4 [61].

### 2.4.5 Underlying events

Due to the complex structure of the protons in the LHC, it is possible to have more than one parton hard-scattering, that is, events for which more than one hard parton-parton interactions occur in the same proton-proton collision [62]. For events with fixed final state invariant masses, the cross section for multiple parton interactions increases with energy, since higher energy interactions probe partons with lower momentum fractions. This means that events with lower invariant masses would receive a bigger impact from multiple hard parton scatterings. This set of events is called Multiple Parton Interactions (MPI). Other effects of the scattering include the beam remnants

---

<sup>12</sup>At this point it should be mentioned that this does not mean that the diquark is to be seen as an elementary particle, but that the interaction treats the diquark as a whole element, instead of acting on its individual particles.

from the proton-proton collisions, which do not take an active part in the initial-state radiation or the hard-scattering process. These effects are often called “underlying events”.

Methods for the simulation of the underlying events are also available and improve the simulation of the events. Jimmy [63], Pythia [50] and Herwig [48, 49] include models for MPI. The physics analyses attempt to reduce the impact of these effects and the methods used for that rely on a good description of them in the Monte Carlo generators.

## 2.5 Beyond the Standard Model

The Standard Model is often seen as an effective model, since there are a set of unanswered questions in it. Open points in the SM include the large top quark mass compared to the other quarks and the hierarchy problem. The hierarchy issue arises because the Higgs boson’s mass is much smaller than the Planck scale, while one would expect that the renormalised mass of the Higgs boson includes a correction that makes it very big. In the SM this could happen if there is a fine-tuning cancellation in the calculation of the renormalised Higgs boson mass, but the reason why this would happen is still an open question. Another unsolved problem is how to incorporate gravity in the Standard Model, since this force is not included in the model so far.

The top quark’s large mass suggests that it might be a window to unknown effects at higher energy. It is also close (that is, much closer than other SM particles) to the electroweak symmetry breaking scale, since the vacuum expectation value for the Higgs field is  $v = 246$  GeV and the top quark mass is  $m_t = 172.5$  GeV, which has lead some researchers to argue that it might be connected to the mechanism of electroweak symmetry breaking [64].

Many models exist which try to extend the Standard Model in such a way that the SM is approximately correct for the phase space region where evidence is available, but they predict new scenarios in situations which still need to be probed. A method for testing the hypothesis on whether these extensions model nature is to extend measurements into regions not probed previously, either with higher energies or increasing precision. The agreement of the SM with experimental results is very good and the existing open questions relate to situations which are still unknown even in the current model’s prediction.

A class of models which extend the SM demand that one extra symmetry is obeyed by the fields, called “supersymmetry”. This extra symmetry demands

a duality between bosons and fermions, creating a list of “superpartners” to the SM particles. For a review of this class of models, consult [65].

Another set of models are the Technicolor models, which demand the existence of new  $SU(3)$  symmetries, similarly to QCD. The new symmetry would create bound states that are seen as the current SM particles. One of these models predicts a new particle which decays to a top-antitop pair and it is called the Topcolor model [66–68]. The electroweak symmetry is broken, in this model, by a bound state and not by the Higgs boson. Alternatives to include gravitation in the Standard Model also exist. They include the Randall-Sundrum (RS) model [69, 70], which predicts an extra dimension in which gravity would propagate. The latter model also includes a particle, called a Kaluza-Klein gluon, that propagates in the extra dimension, which decays into a top-antitop pair.

## 2.6 Summary

A brief review of the main aspects of the Standard Model and how its predictions are done has been made. The Standard Model description is used repeatedly in the physics analyses that follow to describe the well known backgrounds in many analysis, which study Standard Model signals or Beyond the Standard Model physics. The physics analyses are also done repeatedly using the Monte Carlo simulation of events, which includes the matrix element generators, the parton shower simulation, with an appropriate matching between the two, and a hadronisation model.



## Part II

### The experimental setup

# Chapter 3

## The ATLAS experiment

The analyses done in this thesis were performed using data taken by the ATLAS experiment [71] at the Large Hadron Collider [12], therefore a description of the most relevant aspects of this experiment is important. This chapter starts by describing the overall geometry and subdivision of the ATLAS detector followed by some details on its subsystems. This chapter will be used as well to introduce some common notation and conventions.

### 3.1 The ATLAS detector

The Large Hadron Collider (LHC) [12] is a synchrotron located in a tunnel with a 27 km circumference, in the border region between France and Switzerland. The LHC is 100 m below the ground and it collides beams of protons at different points, so that particle detectors can analyse the results of the collisions. Figure 3.1 shows a schematic overview of the LHC. This thesis analyses the measurements from the proton-proton collisions in the ATLAS [71, 72] detector.

ATLAS' systems cover both the barrel of the cylinder and the endcaps, in a structure designed to cover as much of the full  $4\pi$  sr solid angle as possible. A schematic of the ATLAS experiment is shown in Figure 3.2. It is a general purpose detector at the LHC, with many sub-detectors, which measure specific observables of the particles that come out of the collision. Its sub-detectors have cylindrical shapes, with increasing radius, each one encapsulating the smaller ones. It can be divided in four main parts: the Inner Detector; the Magnet System; the Calorimeters and the Muon Spectrometer.

A few conventions used in ATLAS should be mentioned. The coordinate system [72] used in ATLAS is centred at the collision point, in the beam pipe,

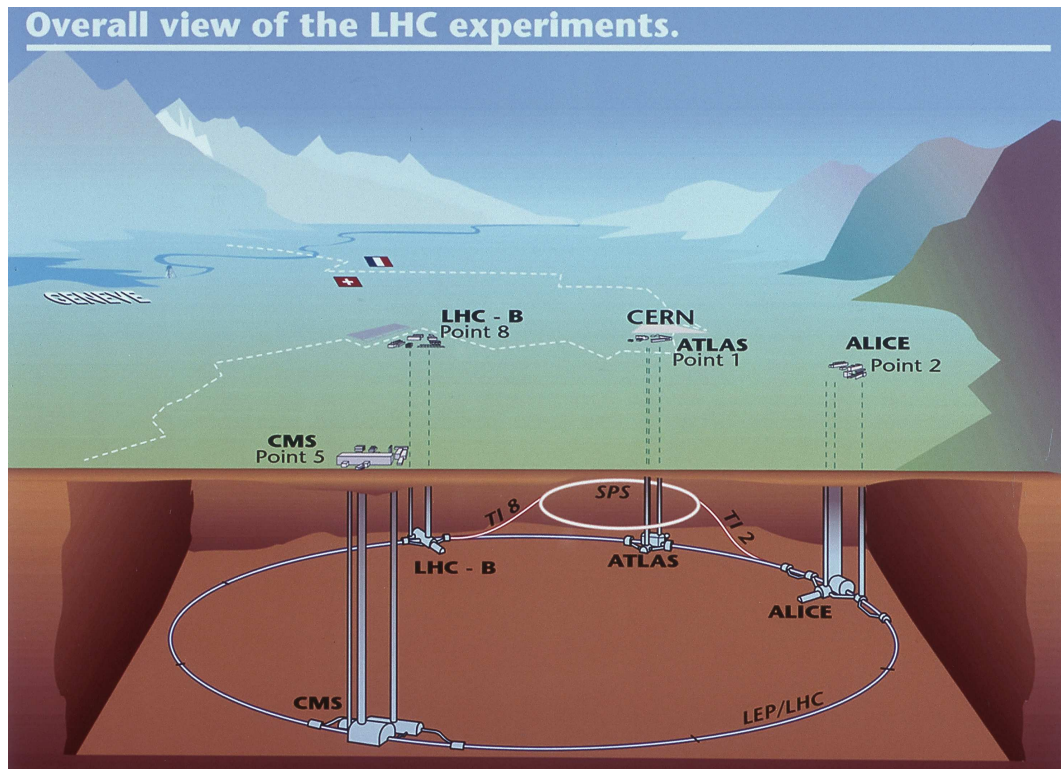


Figure 3.1: Schematic view of the Large Hadron Collider and other particle accelerators with the indication for the experiments built in the LHC ring. All credits to ©CERN.

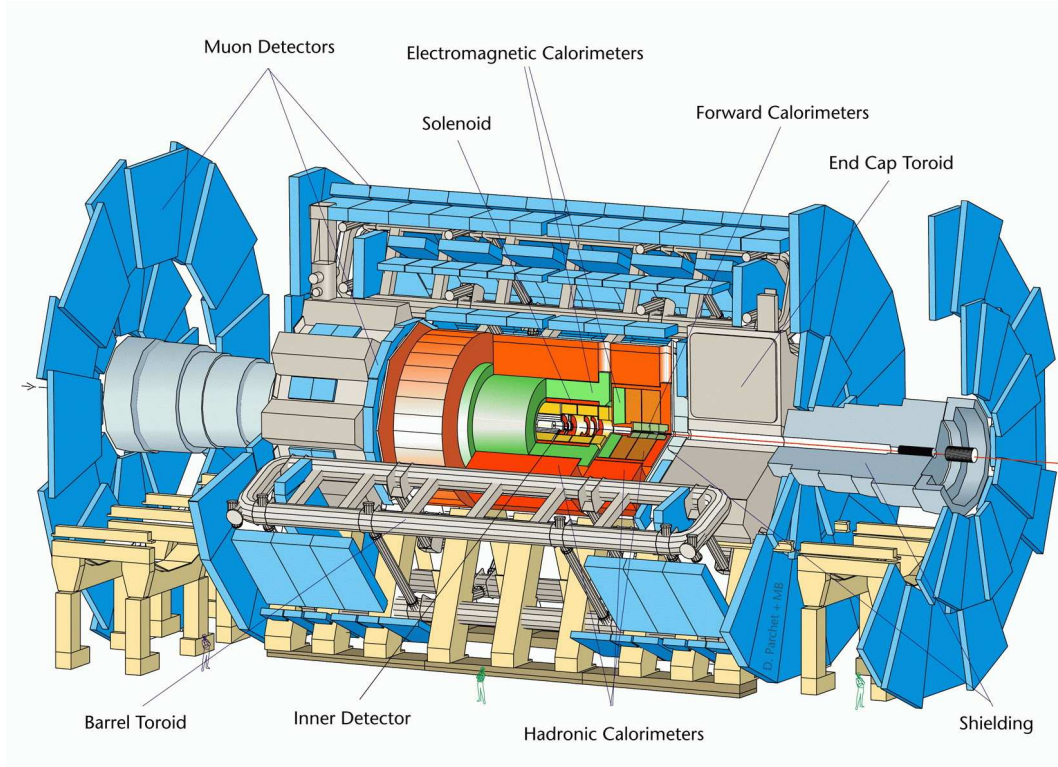


Figure 3.2: A schematic view of the ATLAS experiment. All credits to ©CERN.

with the  $x$ -axis pointing to the center of the LHC tunnel's circumference, the  $y$  axis pointing upwards and the  $z$ -axis chosen so that a right-handed coordinate system is used. The azimuthal angle  $\phi$ , measured in the  $x - y$  plane, and the polar angle  $\theta$ , measured from the positive  $z$  direction, in the  $z - y$  plane, are also used. Frequently, the pseudo-rapidity  $\eta = -\ln(\tan \frac{\theta}{2})$  is used, instead of the angle  $\theta$ .

A measure of the separation between two physical objects detected in ATLAS based only on the direction of their momenta is used in the physics analyses. It can be defined, for two objects  $o_1$  and  $o_2$  with directions, given in the  $\eta - \phi$  plane, by  $(\eta(o_1), \phi(o_1))$  and  $(\eta(o_2), \phi(o_2))$  respectively as:

$$\begin{aligned} \Delta R(o_1, o_2) &\triangleq \sqrt{(\Delta\eta(o_1, o_2))^2 + (\Delta\phi(o_1, o_2))^2} \quad , \text{ where:} & (3.1) \\ \Delta\eta(o_1, o_2) &\triangleq \eta(o_2) - \eta(o_1), \\ \Delta\phi(o_1, o_2) &\triangleq \min(|\phi(o_2) - \phi(o_1)|, 2\pi - |\phi(o_2) - \phi(o_1)|). \end{aligned}$$

It is common to refer to the momentum of particles projected in the  $x - y$

plane, which is called the “transverse momentum” and it is often represented as  $p_T$ , while the magnitude of the three-momentum is referred to as  $p$  and the particle’s energy is referred to as  $E$ . The transverse momentum and the momentum  $p$  can be related by  $p_T = p / \cosh(\eta)$ . A “transverse energy”  $E_T$  is defined by analogy,  $E_T = E / \cosh(\eta)$ .

The ATLAS sub-detectors can be divided in the Inner Detector, the Calorimetry System and the Muon Spectrometer. The next sections will briefly describe these subsystems.

### 3.1.1 Inner Detector

The innermost subdetector of ATLAS is the Inner Detector [11,73] (ID), which is subdivided into the Pixel Detector, the Semi-Conductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

The precision tracking detectors, comprised of the Pixel Detector and the SCT, cover a region of  $|\eta| < 2.5$ . In the barrel region, they are arranged as cylinders around the beam axis, while in the endcap they are arranged as disks perpendicular to the  $z$ -axis. The Inner Detector’s function is to measure tracks generated by charged particle’s interaction with the detectors. Each point in which there is an interaction between the particle and the ID is called a “hit”.

A 2 T solenoidal magnetic field, created by a thin superconducting solenoid surrounding the inner detector, bends the charged particles’ trajectory. An estimate of the trajectory can be done by performing a fit of the hits, and the charge to momentum ratio of the particle can be estimated by calculating the curvature of the track. The direction of the bending also gives the sign of the particles’ charge.

The Pixel Detector has three layers, including one of them at a radius of 4 cm, called  $B$ -layer, which is essential for good vertexing. The basic elements of the Pixel Detector are  $50\ \mu\text{m}$  wide in  $R - \phi$  and  $400\ \mu\text{m}$  long in the  $z$ -axis. The SCT has four cylindrical layers of silicon strips aligned in the azimuthal direction in the barrel, and nine disks in each of the endcaps. In the barrel, the SCT uses eight layers of small-angle strips to measure both coordinates, with one set of strips parallel to the  $z$ -axis which measures the  $R - \phi$  direction. Each silicon microstrip layer is 6.4 cm long and has 768 sensors with a strip pitch of  $80\ \mu\text{m}$ , in the barrel. In the endcaps, the strips are radial. The ID sub-detector with biggest radius is the TRT, which consists of  $\sim 36$  layers of 4 mm diameter straw tubes, with a radiator between them to stimulate Transition Radiation (TR) from electrons. The TRT covers the region of  $|\eta| < 2.0$ .

The tracks are fitted using different algorithms, which aim at a good estimate of the charge-to-momentum ratio, the particle's  $p_T$ , the trajectory and the point of closest approach to the primary vertex in the  $x - y$  plane and in the  $z$ -axis. The distance of the track to the primary vertex in the  $z$ -axis and in the  $x - y$  plane are referred to, respectively, as the longitudinal transverse parameter and the transverse impact parameter (frequently used symbols are  $z_0$  and  $d_0$ ). A set of basic track quality requirements are frequently used in the analyses to demand well-reconstructed tracks with demands on the number of hits and on the impact parameters.

### 3.1.2 Calorimeters

After the Inner Detector, the detectors with bigger radius are, respectively, the Liquid Argon Calorimeter and the Tile Calorimeter [11], which measure the energy of the particles. The objective of the calorimeters is to measure the energy and direction of the particles. The particle interacts with the calorimeter creating a shower of secondary particles. In a sampling calorimeter, such as the ones used in ATLAS, the calorimeter has alternate layers of a material that starts the shower (“absorber material”) and a material that measures the shower’s energy (“sampling material”).

The calorimeters have different structure, depending on whether they are designed to measure electromagnetic showers, produced by particles that interact primarily through the electromagnetic interaction; or hadronic showers, for particles that interact mainly through the strong nuclear force. ATLAS includes an electromagnetic calorimetry system and a hadronic calorimetry system.

The Liquid Argon Calorimeters [74] are used to measure the energy of electromagnetic showers in the barrel and endcap regions and also for measurements of energy in hadronic showers in the endcaps. They are sampling calorimeters with accordion geometry in the barrel to provide symmetry in the measurement as a function of the  $\phi$  coordinate. It is filled with liquid argon cooled by a cryogenic system. Layers of lead and stainless steel are interspaced with liquid argon, with the lead acting as an absorber, giving the initial shower development due to its short radiation length. The secondary electrons create ionisation in the gaps of liquid argon, and the copper electrodes register the signal induced by the ionisation electrons drifting across the gap. The electromagnetic calorimeter also includes a presampler detector, followed by three longitudinal layers of the EM calorimeter, called strip, middle and back layers.

At high energy, most of the electromagnetic shower energy is detected in the middle layer. The strip layer has good discrimination against multiple photon showers, due to its small cells. The presampler detector complements the EM calorimeter with a good estimate of the energy lost in the material before the rest of the calorimetry system. The back layer collects the energy deposited by very high energy electromagnetic showers. The endcap calorimeters consist of the outer and inner wheels, which cover  $|\eta| \in [1.375, 2.5]$  and  $|\eta| \in [2.5, 3.2]$  respectively.

Forward calorimeters are also available in the  $|\eta| \in [3.1, 4.9]$  region. The innermost one is the forward electromagnetic calorimeter, which uses liquid argon as the active material and copper as the passive material. The hadronic forward calorimeters follow it and use tungsten as the passive material.

The Tile Calorimeter [75] is a sampling hadronic calorimeter surrounding the Liquid Argon Calorimeter in the barrel. It uses steel as the absorber material and scintillating plates read out by wavelength shifting (WLS) fibers as the active medium. The optical signals read by the WLS fibers are converted into electric signals by photomultipliers (PMTs). It is designed to absorb hadronic showers in the barrel region of the ATLAS experiment.

### 3.1.3 Muon Spectrometer

The final system is the Muon Spectrometer (MS) [11], which is a tracking device embedded in a toroidal magnetic field that measures the charge to momentum ratio of the muons escaping the calorimeters. The MS' layout is based on the magnetic deflection of muon tracks using a system of three superconducting air-core toroid magnets. A barrel toroid with eight coils surrounding the hadronic calorimeter provides the bending of the muon tracks in the  $|\eta| < 1.0$  region. For  $|\eta| \in [1.4, 2.7]$ , two smaller endcap magnets in both ends of the barrel toroid are used. A combination of the magnetic fields of the barrel and endcap toroids are used in the  $|\eta| \in (1.0, 1.4)$  region, called the transition region.

The Muon Spectrometer reconstructs tracks using three layers of Monitored Drift Tube (MDT) chambers in the  $|\eta| < 2.0$  range, two layers of MDT behind one layer of Cathode Strip Chambers (CSC) in the range  $|\eta| \in [2.0, 2.7]$ . Three layers of Resistive Plate Chamber (RPC) in the  $|\eta| < 1.05$  region and three layers of Thin Gap Chamber (TGC) in the  $|\eta| \in [1.0, 2.4]$  region provide a fast response to select events containing muons.

### 3.1.4 The ATLAS Trigger System

The ATLAS detector systems process data with a very high rate of events, most of which are background processes to most physics analyses. The data acquisition system on ATLAS cannot cope with the high data rates and, even if it could, there are technical and financial constraints on the available permanent data storage facilities which limit the amount of data that can be collected.

Furthermore, the time required to analyse the available data is quite long and the processing power necessary to implement the analyses has many constraints. As a consequence, a trigger system has been developed as part of the ATLAS Data Acquisition System to select only collision events with interest for the ATLAS physics analyses.

The ATLAS Trigger System [76] was developed in a very modular and flexible way, so that it could adapt itself to the physics analyses requirements, selecting events that are most relevant to a set of studies and rejecting most of the common backgrounds. There is a configurable infrastructure designed in a way that can be changed according to the decision of which analyses have priority and which analysis methods are used.

The trigger system has three layers. The first layer is hardware-based and it implements a pre-selection of the events using a coarser granularity of the calorimeters and the muon spectrometer, than the next levels. The second and third levels are implemented in software and analyse data with a finer granularity, using the Inner Detector tracking system as well. The latter is frequently referred to as the High Level Trigger. The three-tier system is designed to work modularly: the Level 1 selects Regions of Interest, in which it detects a particle candidate; next, the Level 2 software processes only the data in that Region of Interest to test the hypothesis that it contains a relevant signal; finally, the third level, or Event Filter, analyses the events accepted by the Level 2, scanning the whole detector with fine granularity. This division not only allows for a modular design, it also reduces the rate of events to be processed in the next layers, so that they can take more time to analyse each event with more complex algorithms. A schematic view of the ATLAS Trigger system is shown in Figure 3.3, with the approximate event rates in each layer.

The first level of triggering, as far as the calorimetry system is concerned, only has access to  $\Delta\eta \times \Delta\phi$  regions of  $0.1 \times 0.1$ , which are called “trigger towers”. Its decision can only be made on the energies deposited in those towers. The Level 2 algorithms can use the high transverse energy deposition in



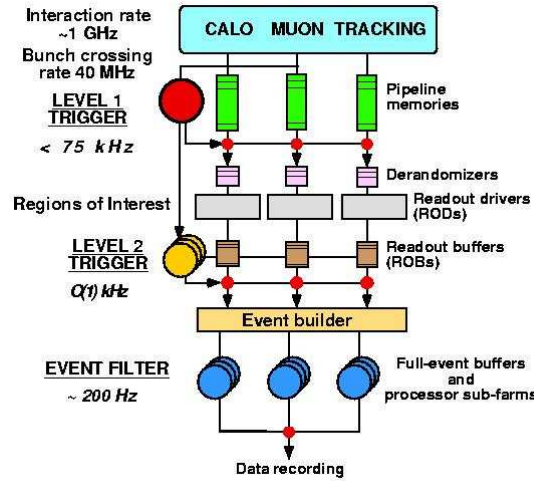


Figure 3.3: A simplified schematic view of the ATLAS Trigger System. Extracted from [77].

the calorimeter as a seed to analyse regions of interest and also take advantage of the Inner Detector and Muon Spectrometer tracks. The Event Filter can also perform a full sliding window search in the event, with access to the full event data, in which the detector is scanned for high energy deposits. The first level of triggering also uses the RPC and the TGC, in the Muon Spectrometer to trigger on muon events with a minimum transverse momentum threshold. the Trigger System also allows for a finer granularity search in the Calorimetry System using the “Level 1.5”, which can be used to seed the Level 2 algorithm instead.

Besides this three layer division, each layer is further separated, depending on its physics goals. This means that if a particular physics object is desired, a specific algorithm will be designed to select only that object in the three layers. In case a muon and an electron are required for the analysis, the system could be configured in a way that the accepted events must have fulfilled the requirements for both the electron and muon selection algorithms. These algorithms are organised and configured using the concept of “trigger chains”.

Each trigger chain has a structure similar to the one shown in Figure 3.4. The chain starts when the hardware-based Level 1 triggers that the event passes the threshold configured for this chain. This is indicated in the “L1 threshold” block in the figure. If the threshold uses the calorimeter information, the Level 1 sums the energy in calorimeter Trigger Towers arranged in a specific way and checks if the sums (there might be more than one criteria) are above or below a threshold. For the electron- and photon-related triggers, for example, the Level 1 trigger looks for a region of four Trigger Towers as a

square in the electromagnetic calorimeter, in which at least one of the four possible two-tower sums (the sum of two towers either vertically or horizontally) of the nearest neighbouring towers pass a pre-defined threshold. Isolation veto thresholds can also be configured for the Trigger Towers around the center four Trigger Towers, as well as for the hadronic calorimeter towers. The algorithm scans all squares with 16 Trigger Towers in the calorimeters. Figure 3.5 shows a representation of all these Trigger Tower sum configurations.

The Level 1 jet trigger scans the calorimeters using elements of four Trigger Tower arranged in a square, summing the electromagnetic to the hadronic trigger towers. The Region of Interest is defined as the square region with four Trigger Towers in  $\eta \times \phi$  space and the sum of the energy on these jet elements is calculated to check if it passes the minimum energy threshold. The window used for this scan can be configured for each chain to have different square sizes in  $\eta \times \phi$  space: 4, 6 or 8 Trigger Tower elements. For regions with 36 Trigger Towers, the Region of Interest can be in four different positions, but for the regions with 8 Trigger Towers in each side, the Region of Interest is required to be in the center of the window, to avoid the possibility of finding two Regions of Interest in the same window. Figure 3.6 shows a representation of the Region of Interest and its possible positions in the window configured.

The Level 1 part of the muon-based trigger chains use the RPC sub-detector in the barrel and the TGC in the endcaps. In the barrel, the RPC is divided in three sectors: RPC1, RPC2 and RPC3, which are composed of two independent detector layers that measure both the pseudo-rapidity and the azimuthal angle of a hit. When a first hit is found in a region of the detector another hit is searched for, in the region defined between the original hit and the interaction point, with an allowed perpendicular width. The allowed width is a parameter that depends on the transverse momentum threshold: the smaller the allowed deviation from a straight line, the higher is the minimum  $p_T$  threshold. Two (three) trigger sectors in coincidence are required for low (high) transverse momentum algorithms.

After a trigger from the configured Level 1 threshold, as shown schematically in Figure 3.4, the High-Level Trigger starts to operate in Level 2, with the information that there was a Level 1 Region of Interest in a certain pseudo-rapidity and azimuthal angle. The general organisational structure of the Level 2 and the Event Filter are similar in that they are sub-divided in two types of elementary blocks: the Feature Extraction algorithm and the Hypothesis testing algorithm. The Feature Extraction algorithms calculate relevant vari-

ables using the available Inner Detector, calorimeter and Muon Spectrometer information. These variables are stored and can be refined by the next Feature Extraction algorithm in the sequence or used by the Hypothesis testing algorithm to demand that certain logical criteria are fulfilled. The Hypothesis testing algorithm simply decides whether to keep the event (if one or all of the criteria are fulfilled), or to reject it. If the event is accepted at the Level 2, it proceeds to be examined by the Event Filter, using the same organisation for Feature Extraction and Hypothesis testing algorithms. The event is finally accepted if it passes the Event Filter hypothesis testing. The algorithms used in the Level 2 and Event Filter Feature Extraction algorithms depend on which physics object the chain should accept. Consult [11] for details.

## 3.2 Multiple interactions in ATLAS

Each bunch crossing analysed by ATLAS includes a beam of protons in both  $+z$  and  $-z$  directions. More than one interaction is expected to happen and the measurement of the distribution of the mean number of interactions per bunch crossing is shown in Figure 3.7 for 2011 and 2012 data. This effect is often referred to as “pile up” and it has a significant impact in the physics analyses. One example of the effects of pile up is that extra particles are produced in the final state as a result of pile up interactions that could be confused as coming from the interaction under analysis.

The Monte Carlo simulations do not perfectly simulate the shape of the  $\langle \mu \rangle$  distribution extracted in data, which leads to a discrepancy when comparing data and simulation. This can be fixed in the physics analyses by weighting the events by the ratio of the data to simulation  $\langle \mu \rangle$  distributions. More details about this reweighting procedure are given in Section 5.6. The  $\langle \mu \rangle$  value is estimated as an average number of interactions over the time period of a “luminosity block” in ATLAS. The duration of a luminosity block is set by the Data Acquisition System of ATLAS, which is of approximately two minutes.

## 3.3 Electron reconstruction and identification

Electrons can be initially identified by the Trigger System [76] through a selection chain in all three levels of triggering. After a trigger selection, the offline algorithms reconstruct the electron four-momentum using the electromagnetic

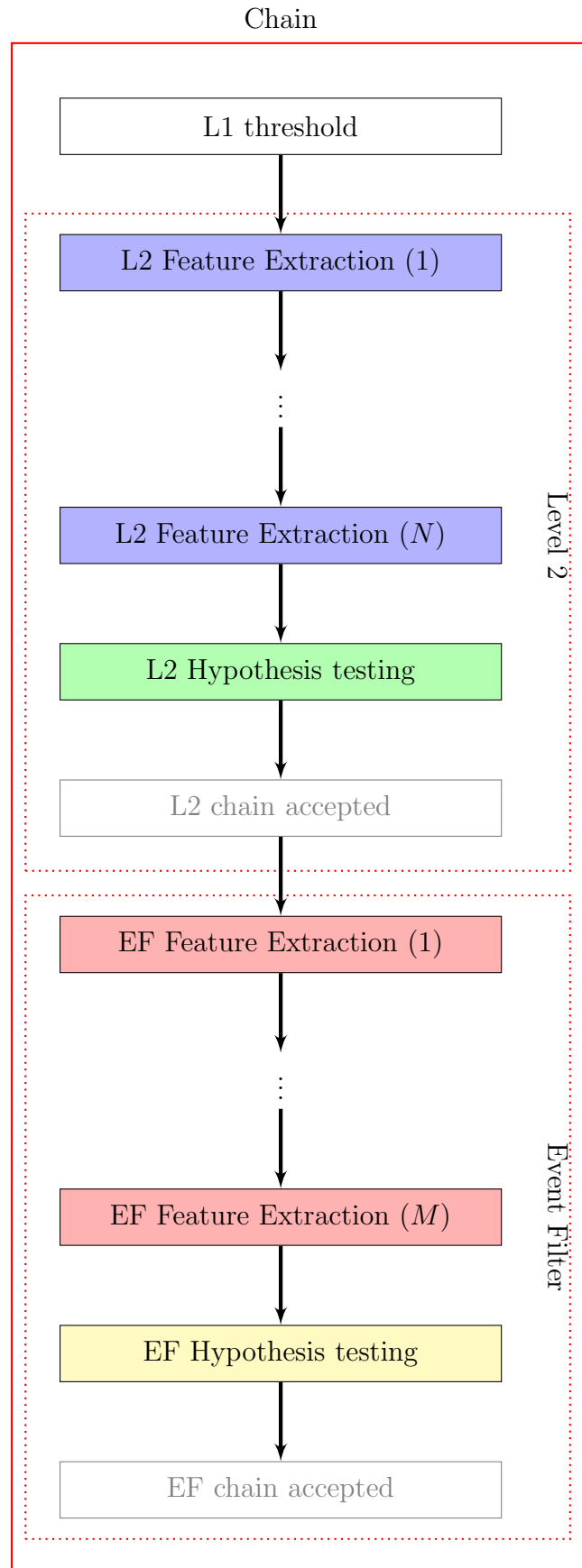


Figure 3.4: Simplified schematic that shows the structure of the trigger chains.

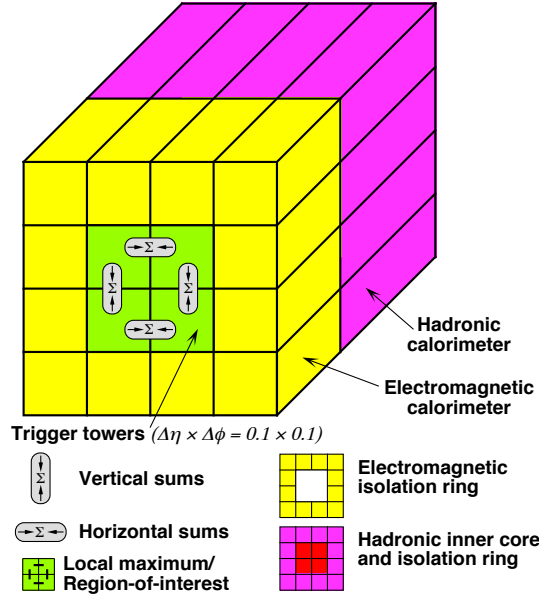


Figure 3.5: Schematic representation of the Trigger Towers used to calculate the electron/photon-related Level 1 trigger threshold sums. The core of  $2 \times 2$  trigger towers in the electromagnetic calorimeter is required to contain the sum of two Trigger Towers horizontally or vertically that satisfy the minimum threshold. Isolation veto using the ring of cells around the center ones and the hadronic calorimeter energy sums can also be implemented in some chains. Extracted from [11].

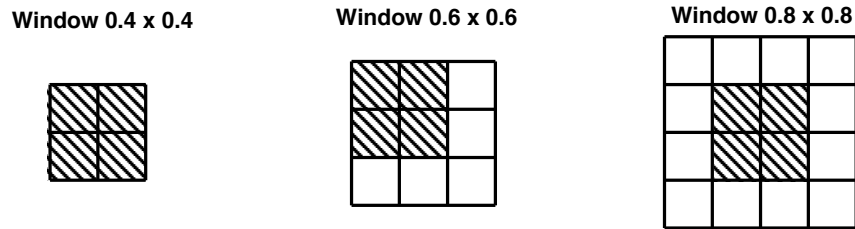


Figure 3.6: Schematic representation of the Trigger Tower sum configuration for the jet-related triggers at Level 1. Extracted from [11]. The jet trigger algorithms are based on jet elements which have the size of  $2 \times 2$  Trigger Towers. The Region of Interest is shaded. For scans using  $6 \times 6$  windows, there are four possible windows containing a Region of Interest, but in the  $8 \times 8$  case, the Region of Interest is required to be in the center position, to avoid the possibility of two jets in a single window.

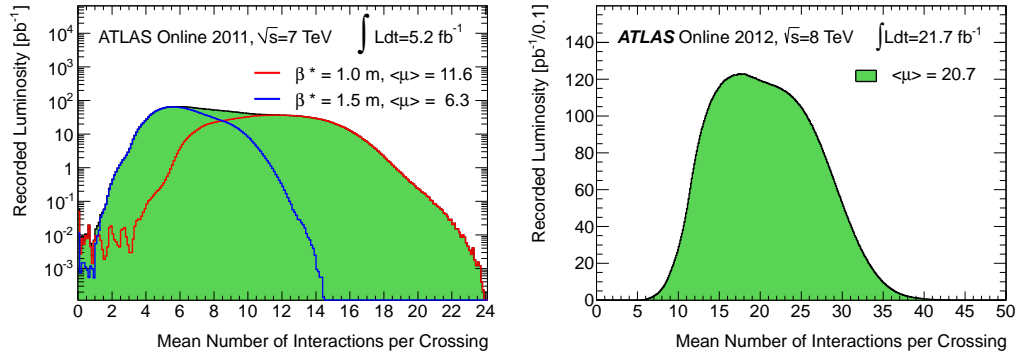


Figure 3.7: The mean number of proton-proton interactions per bunch crossing in ATLAS is shown for the data taking in 2011 (left) and 2012 (right). For 2011, the set up after the Technical Stop in September (with  $\beta^* = 1.0$  m) is shown in red and the set up before it is shown in blue (with  $\beta^* = 1.5$  m). ATLAS performance public plot not produced by the author. More information about the measurement can be found in [78]. Entries in  $\langle \mu \rangle \sim 0$  arise from pilot bunches that were present in many early LHC fills.

calorimeter clusters as seeds<sup>1</sup>, which are matched to ID tracks at a later stage to identify and reconstruct the electron’s four-momentum [79]. Identification algorithms based on the track’s characteristics and the energy deposition in the calorimeter are used. A full account of the methods used for the trigger, reconstruction and identification of electrons in ATLAS can be found in [79]. It is important to highlight the calibration and correction factors applied to correct for the difference in behaviour of these algorithms in real data and in Monte Carlo simulation. These corrections are calculated based on statistical analyses in simulation and real data and the methods used to extract them will be mentioned here, since these corrections are used at a later stage, in the analyses chapters.

The energy scale and resolution of the electrons can be calculated in real data, using  $Z$ -boson and  $J/\Psi$  decays into pairs of electrons, which are required to have a minimum transverse energy of 20 GeV for the  $Z$ -boson decays and 5 GeV for the  $J/\Psi$  decays and to pass an electron trigger requirement. The invariant mass of the electron pairs is required to be in a window around the masses of the  $Z$ -boson (80 GeV - 100 GeV) or the  $J/\Psi$  (2.5 GeV - 3.5 GeV) and the leptons are required to have opposite charges. The amount of background is estimated and subtracted. With these events, the relation between the true electron energy,  $E^{\text{true}}$ , and the measured energy,  $E^{\text{measured}}$ , in a region  $i$  of the

<sup>1</sup>Photons are also selected based on similar requirements on the electromagnetic calorimeter clusters. Photons would have no ID tracks, though.

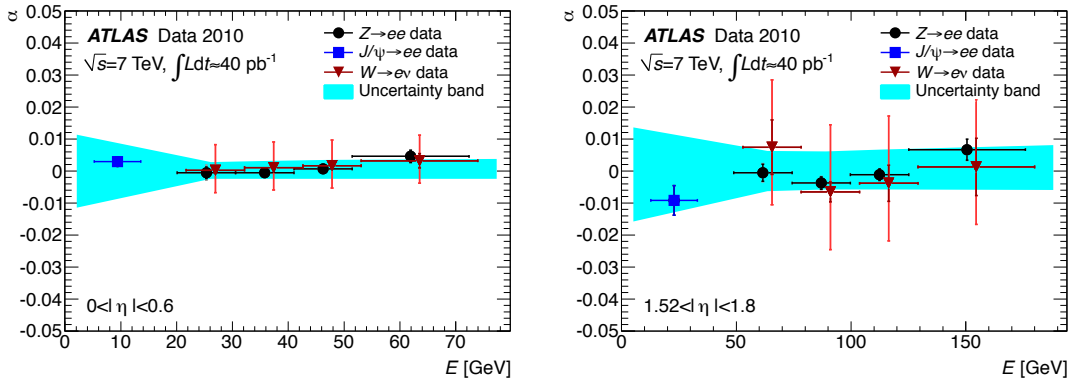


Figure 3.8: Results of the measurement of the electron energy scale in  $Z \rightarrow e^+e^-$  decays and in  $J/\Psi \rightarrow e^+e^-$  decays in ATLAS 2010 data, for  $|\eta| < 0.6$  (left) and  $1.53 < |\eta| < 1.8$  (right). Extracted from [79].

detector is given by:

$$E^{\text{measured}} = E^{\text{true}}(1 - \alpha_i). \quad (3.2)$$

An unbinned likelihood function is maximised [79] to obtain the values of  $\alpha_i$  for every region of the detector. The results of this measurement for 2010 ATLAS data are shown in Figure 3.8. Cross-checks of these results were done using other methods, which rely on  $W^\pm \rightarrow e^\pm \nu_e$  decays. The  $\alpha_i$  terms in the formula above are used to correct the simulated electron energy scale so that it behaves similarly to data in the analyses that follow and the systematic uncertainties associated with its measurement [79] are also taken into account.

The electron fractional energy resolution is parametrised as:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (3.3)$$

where  $a$ ,  $b$  and  $c$  are empirical parameters, and the parameter  $c$  is frequently referred to as the “constant term” in the fractional energy resolution parametrisation. Figure 3.9 shows a fit in data and the Monte Carlo simulation prediction used to derive the energy resolution parameters. Several sources of systematic uncertainties are investigated and more details can be found in [79]. A smearing of the electron kinematics is applied in simulation for the physics analyses to correct for the difference in the energy resolution in data and simulation.

The electrons are reconstructed, furthermore, demanding that they satisfy the trigger requirement, a set of track- and calorimeter-related cuts, a set of reconstruction requirements and, in some analyses, an isolation requirement,

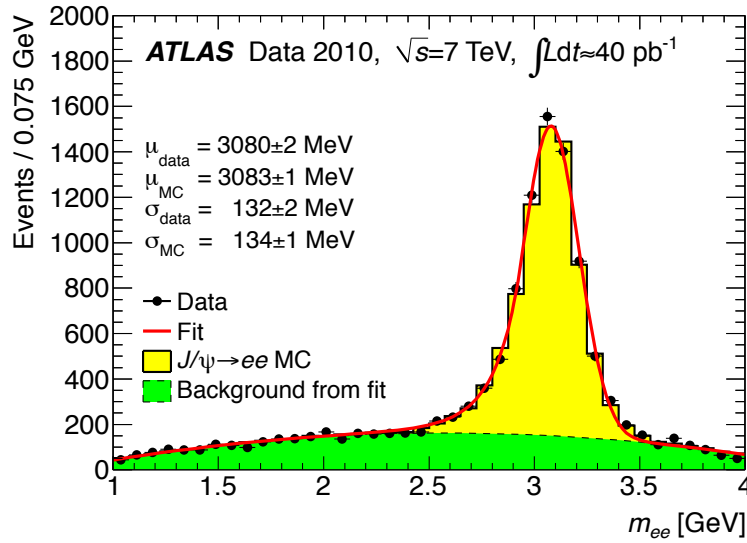


Figure 3.9: Fit of data and simulation for the electron energy resolution estimate from  $J/\Psi \rightarrow e^+e^-$  decays using ATLAS 2010 data. Extracted from [79].

which demands that there are no other tracks or energy deposits around the particle. A final set of corrections applied in simulation are related to a difference in electron selection and reconstruction efficiencies in simulation and data. They are applied in the physics analysis by weighting the events by the ratio of efficiencies in data and simulation. These corrections are frequently called “scale factors” in this document. To calculate these corrections, though, it is necessary to measure each of these efficiencies in simulation and data. That was done, for 2010 ATLAS data in [79]. When applying the weights calculated with the derived scale factors in the physics analyses, the uncertainties are taken into account, by varying the weights accordingly.

The efficiency measurements are done using the Tag And Probe method, which studies decays into pairs of particles from real data  $Z$ -boson,  $J/\Psi$  or  $W$ -boson decays. This method aims at selecting a sample of “probe” electrons using selection cuts which are called “tag” requirements, on another physical object, which is called the “tag” object. A selection can then be applied in the “probe” electron to investigate its efficiency. In  $Z \rightarrow e^+e^-$  and  $J/\Psi \rightarrow e^+e^-$  events, one of the electrons is used as a tag, while in the  $W^\pm \rightarrow e^\pm\nu_e$  events, high missing transverse energy is used as a tag. The  $Z$ -boson, the  $J/\Psi$  or the  $W$ -boson masses can be used as constraints to reduce the background contribution in these analyses, which are subtracted.

In the Tag And Probe method, the tag selection is looser than the probe one. A set  $N_T$  of events is chosen so that at least one electron passes the tag requirement. In this set, a subset  $N_{T\&P}$  of events is required to satisfy the



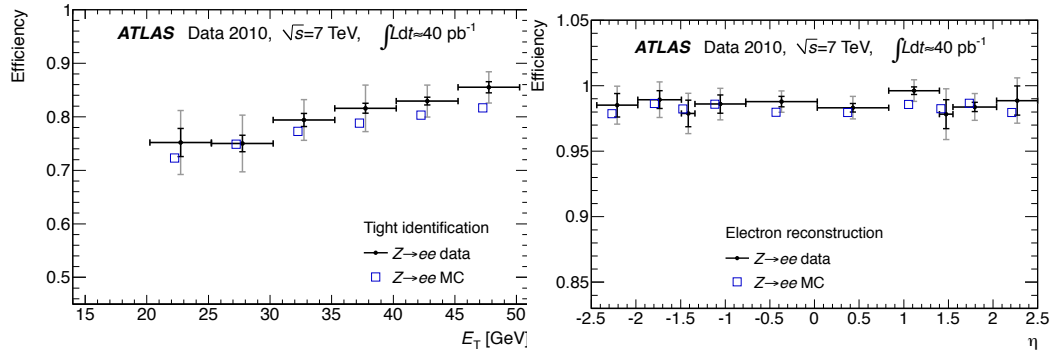


Figure 3.10: Efficiency measurement results using the Tag And Probe method in  $Z \rightarrow e^+e^-$  decays in 2010 ATLAS data for the electron identification (left) and the electron reconstruction efficiencies. Extracted from [79].

probe selection as well. The ratio  $N_{T\&P}/N_T$  is used as a measure of the probe selection efficiency. More details about these efficiency measurements can be found in [79]. Figure 3.10 summarises a few efficiency measurement results for the electron identification and for the electron reconstruction.

### 3.4 Muon reconstruction

The muons are identified initially by the Level 1 trigger [80] using the Muon Spectrometer and they are further analysed by the next trigger levels using the Inner Detector tracks. The events selected by the muon trigger use the full resolution of the detector in the Muon Spectrometer and the Inner Detector to reconstruct the muon’s momentum and charge. Muon events reconstructed only by the Muon Spectrometer, with no match to the Inner Detector track, are called “Standalone Muons”, while muons which have an Inner Detector track that can be associated to straight track segments in the Muon Spectrometer are called “Segment Tagged Muons (ST)”. The muons used in the analyses have a track reconstruction performed separately in the Inner Detector and the Muon Spectrometer, which can be used to form a combined track. The latter are called “Combined Muons (CB)”. Due to differences in the muon momentum resolution in data and simulation, the physics analyses smear the simulated muon momenta so that the corrected simulation resolution matches data. However, to implement this, the resolution in simulation and in data must be measured.

The reconstructed muon tracks have a resolution [81] in data and simulation which can be parametrised, in the Muon Spectrometer (MS), for a given

pseudo-rapidity, by:

$$\frac{\sigma^{MS}(p)}{p} = \frac{p_0^{MS}}{p_T} \oplus p_1^{MS} \oplus p_2^{MS} p_T, \quad (3.4)$$

where  $p$  is the reconstructed momentum,  $p_T$  is the reconstructed transverse momentum,  $\sigma^{MS}$  is the resolution in the MS and  $p_0^{MS}$ ,  $p_1^{MS}$  and  $p_2^{MS}$  are coefficients related to the energy loss in the calorimeter material, multiple scattering and intrinsic resolution terms respectively. In the Inner Detector, the parametrisation can be done similarly, but for the central part of the detector, in  $|\eta| < 1.9$ , the energy loss term can be dropped:

$$\frac{\sigma^{ID}(p)}{p} = p_1^{ID} \oplus p_2^{ID} p_T, \quad (3.5)$$

where the resolution is given by  $\sigma^{ID}$ . For  $|\eta| \geq 1.9$ , the worsening of the resolution due to the edge of the TRT fiducial volume can be parametrised as:

$$\frac{\sigma^{ID}(p)}{p} = p_1^{ID} \oplus p_2^{ID} p_T \frac{1}{\tan^2(\theta)}, \quad (3.6)$$

where  $\theta$  is the angle measured in the  $y - z$  plane for the momentum direction. The muon resolution is measured by analysing  $Z$ -boson decays into muon pairs and  $W$ -boson decays into a muon and a neutrino. The width of the  $Z$ -boson invariant mass reconstructed in selected events after background subtraction [81] is related to the muon resolution. The  $W^\pm \rightarrow \mu^\pm \nu_\mu$  decays can also be used to calculate the relative difference in momentum predicted by the MS and the ID, which should have a mean at zero and its width probes the quadratic sum of the resolutions in the ID and the MS. The resolution for the ID and the MS have been measured [81] in different pseudo-rapidity ranges and they are shown in Figure 3.11.

From the measurements of the resolution in data and simulation, and a fit according to the resolution parametrisation, a correction strategy can be devised for the MS and ID tracks separately, which are then combined in a single combined muon correction, which is applied in simulation. The Muon Spectrometer tracks are corrected in simulation according to:

$$p'_T(MS) = p_T(MS)(1 + \Delta(MS)), \quad (3.7)$$

where  $p'_T(MS)$  is the corrected transverse momentum of the Muon Spectrometer track,  $p_T(MS)$  is the uncorrected transverse momentum of the MS tracks

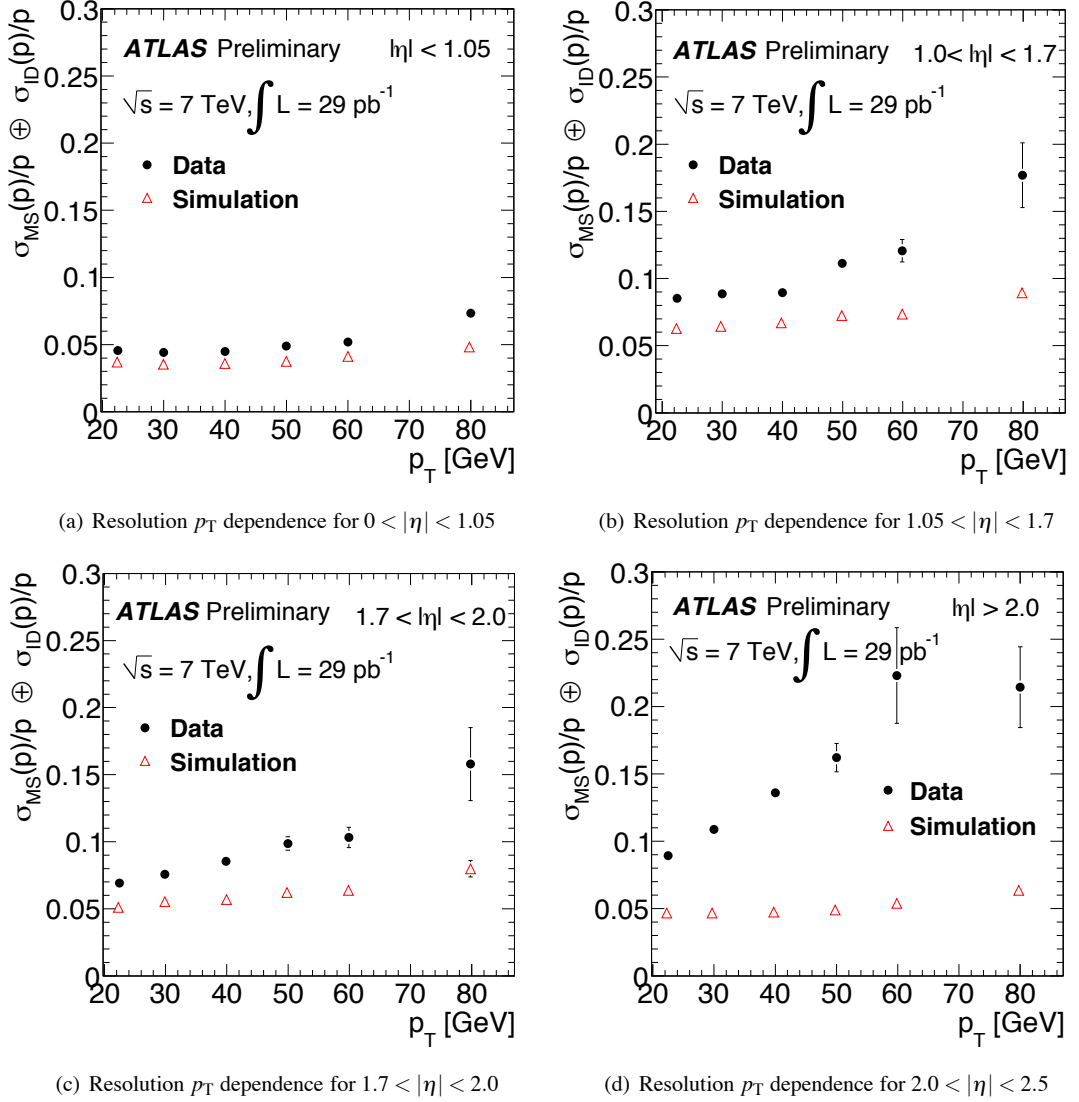


Figure 3.11: Sum in quadrature of the Muon Spectrometer and the Inner Detector muon resolutions as a function of the transverse momentum in four pseudo-rapidity regions using  $W \rightarrow \mu\nu$  events in ATLAS 2010 data. This is the result of a preliminary analysis, on which there were shortcomings in the simulation of intrinsic resolution and module misalignment [81]. Extracted from [81].

and  $\Delta(MS)$  is extracted from:

$$\Delta(MS) = f(0, 1)\Delta p_1^{MS} + f(0, 1)\Delta p_2^{MS} p_T(MS), \quad (3.8)$$

where  $f(0, 1)$  is a random number from a Gaussian sample with mean zero and variance one and  $p_i^{MS}$  are the fit parameters. For the Inner Detector tracks, the correction is applied in a similar way, but considering another parametrisation for the  $\Delta(ID)$  correction:

$$\begin{aligned} \Delta(ID) &= f(0, 1)\Delta p_2^{ID} p_T, \text{ for } |\eta| < 1.9, \\ \Delta(ID) &= f(0, 1)\Delta p_2^{ID} p_T / \tan^2(\theta), \text{ for } |\eta| \geq 1.9, \end{aligned} \quad (3.9)$$

where the  $p_i^{ID}$  parameters are the fit parameters for the ID tracks. These resolution corrections can be combined for a combined muon correction using [81]:

$$p'_T(CB) = p_T(CB) \left[ 1 + \frac{\frac{\Delta(MS)}{\sigma^2(MS)} + \frac{\Delta(ID)}{\sigma^2(ID)}}{\frac{1}{\sigma^2(MS)} + \frac{1}{\sigma^2(ID)}} \right], \quad (3.10)$$

where  $\sigma(MS)$  and  $\sigma(ID)$  are the parametrised muon resolutions for the MS and the ID, respectively, and  $p'_T(CB)$  is the corrected combined muon transverse momentum, while  $p_T(CB)$  is the combined muon transverse momentum before the correction. Detailed information on how the fit is performed can be found in [81]. For the purposes of the physics analyses presented in this document, the corrections above are the ones used to smear muons in simulation, to match the data resolution. The uncertainties of the resolution measurements are taken into account in the physics analysis, by shifting the MS and ID resolutions separately, recalculating the smeared combined momentum, and estimating its effect in the analysis.

Besides the muon smearing correction, the efficiencies for muon identification and reconstruction also have a different behaviour in data and simulation, which is corrected in the physics analysis, by weighting the simulation events using the efficiency ratio between data and simulation. Both the Inner Detector reconstruction efficiency, the Muon Spectrometer reconstruction efficiency, the MS to ID track matching efficiency and the muon isolation efficiencies must be taken into account. The efficiency measurement is done using  $Z$ -boson or  $J/\Psi$  decays into pairs of muons in the Tag And Probe method described in the previous section. Figure 3.12 shows the efficiency measured in data and simulation using  $Z$ -boson decays for the 2010 data. More details on the method

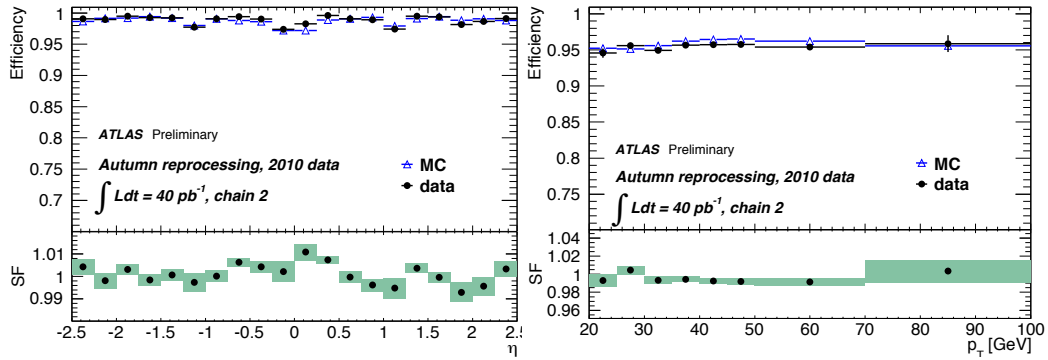


Figure 3.12: Muon reconstruction efficiency not considering the isolation requirement, measured using  $Z$ -boson decays into pairs of muons. In the left figure, the Inner Detector reconstruction efficiency is shown. In the right figure, the efficiency of reconstructing Combined Muons, relative to the Inner Detector efficiency is shown. This was done with 2010 ATLAS data and it was extracted from [80].

and other efficiency results are available in [80].

### 3.5 Jet algorithms

Jets are an important ingredient in the physics analyses shown in this document. They are detected in the experiment as a collection of nearby clusters in the calorimeter and tracks for charged particles in the tracking chambers. Different algorithms can be used to reconstruct a four-momentum which would be related to the particle that initiated the shower and lead to the jet formation, such as the Anti- $k_t$  algorithm [82], or the  $k_t$  algorithm [83]. Both algorithms iteratively combine the momenta of pairs of clusters, if they satisfy a set of criteria until a single jet four-momentum is calculated. They rely on an  $R$  parameter, which provides a relative measure in  $\eta \times \phi$  space for the distances between the jet constituent elements. In the analyses performed in this document, the  $R$  parameter used was either  $R = 0.4$ , or  $R = 1.0$ . In all analyses, the FastJet software [84] is used to implement the jet algorithm used.

The main jet algorithm used in this document is the Anti- $k_t$  algorithm, which starts from a set of four-momenta information, which can be topoclusters, MC meta-data from the simulation or tracks, and calculates, for each pair of four-momenta with transverse momenta  $k_{tj}$  and  $k_{ti}$ :

$$d_{ij} = \min(k_{tj}^{2p}, k_{ti}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.11)$$

where  $\Delta R_{ij} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  is the distance in the  $\eta \times \phi$  plane between the two four-momenta.  $R$  is a parameter of the algorithm and  $p = -1$  for the Anti- $k_t$  algorithm<sup>2</sup>. For each four-momentum with transverse momentum  $k_{tk}$  the algorithm calculates:

$$d_{kB} = k_{tk}^{2p}. \quad (3.12)$$

It groups each pair of elements  $i$  and  $j$ , summing their four-momentum, into a new element if  $d_{ij} < d_{iB}$  and  $d_{ij} < d_{jB}$ , otherwise it looks for a new pair of elements to try and merge. After all combinations are done to group elements in the jet, the algorithm starts over, trying to group new pairs after the merging of some elements.

For jets with large- $R$ , an interesting observable can be calculated by studying its substructure. One may recluster the jet’s constituents with the  $k_t$  algorithm [82, 83]. The last merging step of the procedure previously described defines a  $d_{ij}$  value called the *first splitting scale*  $\sqrt{d_{12}}$ . This observable is used in the following analyses to select jets generated by heavy particle decays (see Section 6.4).

Jets which are built using the calorimeter information as an input are called “calorimeter jets”. More specifically, jets in this document use topological calorimeter clusters (“topo-clusters”) [85], which are built from topologically connected calorimeter cells with significant energy above a noise threshold. The topo-clusters are initially reconstructed at the EM scale [86], which measures the energy of particles produced in electromagnetic showers in the calorimeter. The clusters can be calibrated using a Local Cluster Weighting (LCW) method, which improves the resolution, correcting for fluctuations in the calorimeters [87]. Furthermore, an *in situ* calibration of jets is applied to correct the jet energy scale in simulation, so that it matches data [87]. Jets can also be built from particles’ four-momenta in the shower generated in simulation, without any detector simulation and these jets are called “particle jets” or “truth jets”.

Different techniques can be used to measure the jet energy scale in data and simulation. One method selects only events with a photon or a  $Z$ -boson and one extra jet. In this method, called “direct balance” (DB), due to the momentum conservation in the event, the jet should recoil with a momentum opposite to the  $Z$ -boson or photon, which provides a measurement of the jet energy relative to the  $Z$ -boson or photon energies [87]. Another method,

---

<sup>2</sup>The  $k_t$  algorithm follows the same definition for what follows, but it has  $p = 1$  instead.

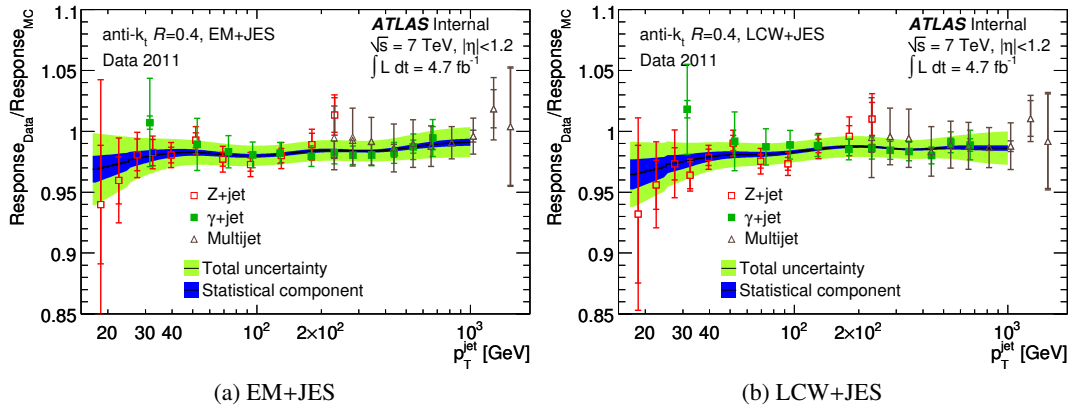


Figure 3.13: Ratio of the jet energy scale in data and simulation for Anti- $k_t$   $R = 0.4$  jets built using the EM scale (left) or using the LCW method (right) for 2011 ATLAS data. Extracted from [87].

called Missing transverse momentum Projection Fraction (MPF) relies on the transverse momentum balance between a photon and a hadronic recoil, which has its transverse momentum calculated using topo-clusters. A third method selects events with low transverse momentum jets recoiling against a high  $p_T$  jet and it can be used to calibrate high  $p_T$  jets, based on a calibration done on low  $p_T$  jets using the direct balance method. The ratio of the jet energy scale measurements in data and simulation on Anti- $k_t$  jets with  $R = 0.4$ , at EM scale or with the LCW method, is shown in Figure 3.13, with all its uncertainty bands. The *in situ* calibration is applied in the physics analyses that follow and the jet energy scale uncertainty is varied in the analyses to quantify its impact. Similar studies were done for jets with  $R = 1.0$  parameter and the results can be consulted in [88]. More information with other calibration methods can be found in [85] as well.

Another important effect, which is taken into account in the physics analyses is the jet reconstruction efficiency, which measures the efficiency with which truth jets are reconstructed in calorimeter jets using simulation and a Tag And Probe-based method in data [85]. The uncertainty on the jet reconstruction efficiency is also taken into account in the physics analyses. The jet energy resolution was also measured [85, 89] and the jets were smeared when calculating the systematic uncertainty in the physics analyses.

A quantity called “jet vertex fraction” (JVF) is calculated for each jet and a selection cut applied on it can be used to reduce the impact of multiple particle interactions in the analyses. The JVF is calculated as the fraction of the jets’ tracks’  $p_T$  scalar sum for tracks that can be matched to the primary vertex. The primary vertex (PV) is calculated as the vertex associated with the

highest sum of tracks' squared transverse momenta ( $\sum p_{t,\text{track}}^2$ ) for all tracks associated with that vertex.

The efficiency of the jet vertex fraction selection requirement is measured in  $Z + 1$  jet events (in which the  $Z$  boson decays into a pair of leptons), by selecting events with jets back-to-back to the reconstructed  $Z$  boson, similarly to the jet energy scale measurement. The uncertainty on this efficiency measurement was calculated by varying the selection requirements for the  $Z + 1$  jet events. A scale factor was calculated as the ratio of efficiencies of the JVF selection in data and simulation with its appropriate uncertainty, so that the simulation can be corrected in the physics analyses.

### 3.6 *b*-tagging algorithms

Although reconstructing the four-momentum of jets is important, there are many ways in which they could be produced, from the hadronisation of  $u$ ,  $d$  or  $s$  quarks, to a decay of high momentum particles, such as the Higgs boson, or the top quark, which have their decay products overlapping when detected. Identifying the mechanism in which the jet was generated is very important in some analyses, for reducing the background contribution. The top quark, for example, decays almost exclusively to a  $W$ -boson and a  $b$ -quark, while the latter always generates a jet. Identifying a “ $b$ -jet” is, therefore, important to select events which contain top quarks and separate them from the backgrounds, which contain jets generated from  $u$ ,  $d$ ,  $s$ , or  $c$  quarks<sup>3</sup>. The  $b$ -quarks hadronise into a hadron containing a  $b$ -quark, which has a relatively long lifetime of  $\sim 1.5$  ps [7] in the process of generating a  $b$ -jet. This long lifetime can be exploited to identify “ $b$ -jets”.

Many methods have been developed to identify whether a jet is a  $b$ -jet, or another type of jet. Jets coming from a  $c$ -quark hadronisation are called “ $c$ -jets”, while jets which originated from  $u$ ,  $d$ ,  $s$  quarks or gluons are jointly called “light-jets”. In general the algorithms used to identify  $b$ -jets have a weight as an output, on which a cut can be made, depending on the desired efficiency and mistag rate (that is, the probability of selecting jets, although the jets are not generated by a  $b$ -quark). The algorithm used in the physics analyses in this document has been configured to have a 70% efficiency over a broad transverse momentum range and it is called “MV1”. It uses a neural

---

<sup>3</sup>As it will be seen in the physics analyses Sections 5.2 and 5.3, the backgrounds include  $W$ -boson production with extra jets,  $Z$ -boson production with extra jets and QCD multi-jets, which contain a smaller fraction of  $b$ -jets than the top quark production.



network that combines the output weight of other algorithms, named “IP3D”, “SV1” and “JetFitterCombNN”, which are described in [90,91].

In the *b*-tagging algorithms, the jets are associated with tracks, requiring that all tracks associated with a given jet have a maximum  $\Delta R(\text{jet}, \text{track})$ . The maximum  $\Delta R(\text{jet}, \text{track})$  required varies with the transverse momentum of the jet, so that high momentum jets have a smaller cone size and are more collimated. The tracks are then selected according to some quality cuts criteria (see [91] for more details). Their four-momenta and the position of their perigee to the primary vertex are used to compute a weight in each algorithm. These weights are expected to peak in a specific number if the jet is a *b*-jet, so that a selection requirement can be designed based on it.

The IP3D algorithm calculates a likelihood function for signal (*b*-jets) and backgrounds (*c*-jets and light-jets) in simulation for the significance of the longitudinal impact parameter,  $S(z_0) = z_0/\sigma(z_0)$ , and the significance of the transverse impact parameter,  $S(d_0) = d_0/\sigma(d_0)$ , of each track associated with the jet. A likelihood ratio is calculated for the jet, by multiplying each of the track likelihood functions. The significances for the impact parameters are signed, based on whether the track intercepts the jet cone axis or not <sup>4</sup>. For *b*-jets, the likelihood function for these signed significances is asymmetric, while it is expected to be approximately Gaussian for the light-jets.

The SV1 algorithm, calculates the position of the secondary vertex (SV) and associates it to the jet tracks. It uses a likelihood ratio technique based on a few variables calculated based on the SV position: the invariant mass of the sum of the tracks associated to the SV, the ratio of the sum of energies of tracks associated with the SV and the sum of energies of all tracks associated with the jet, and the number of vertices in the event. The former two variables are combined in a two-dimensional likelihood, while the latter variable is used to build a one-dimensional distribution. Other information is also used, such as the distance between the jet axis and the line that goes through the primary and secondary vertices. More information about this tagger can be found in [91]. The JetCombFitterNN algorithm relies on the topology of the hadron decays inside the jet and more about it can be read in [90–92].

Different methods can be used to measure the efficiency of the *b*-tagging algorithms in data, which are given in details in [90,92]. Figure 3.14 shows the efficiency of the MV1 algorithm, calculated in simulation and data using

---

<sup>4</sup>The convention used for the sign association is not essential for the algorithm, as long as it is used consistently.

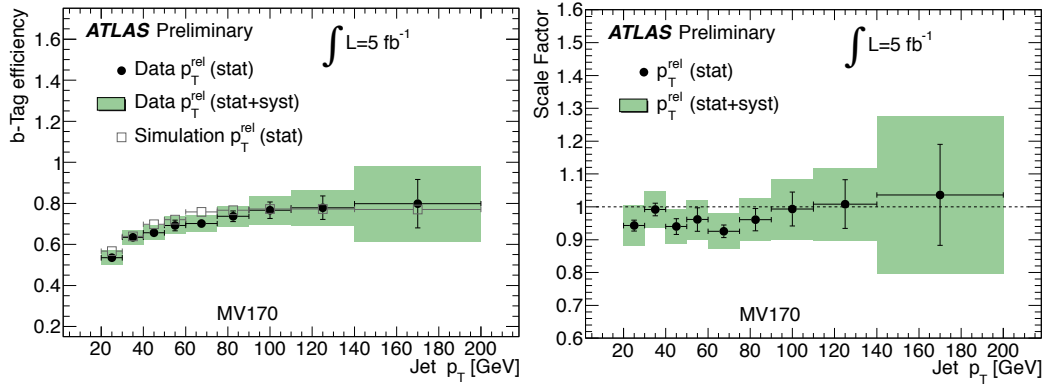


Figure 3.14: The efficiency in data and simulation (left) and their ratio (right) for the MV1 b-tagging algorithm in its 70% efficiency working point, calculated using ATLAS 2011 data and the  $p_T^{\text{rel}}$  method [90, 92]. Extracted from [92].

the  $p_T^{\text{rel}}$  method. This method exploits the semileptonic decay of the  $B$ -hadron to a muon and it reconstructs the momentum of the muon transverse to the combined muon and jet axis. It builds a template for this distribution in  $b$ -jets,  $c$ -jets and light-jets and these are used in a fit in data to obtain the number of  $b$ -jets before and after the  $b$ -tagging requirement. More details about this method can be found in [90, 92].

### 3.7 Missing transverse energy reconstruction

An important component for physics analyses is the measurement of the missing transverse energy in the detector, which calculates the imbalance of the total transverse momentum after the collision. The source of the imbalance is related to particles which are not detected in ATLAS. In the Standard Model, the undetected particles are neutrinos, but other models include different new particles which also would not be detected easily and they could be a source of missing transverse energy.

The measurement of the missing transverse energy,  $E_T^{\text{miss}}$ , is separated in two components: one uses the calorimeter information and the other one uses the Muon Spectrometer information. The calorimeter term uses the calibrated calorimeter cells according to the reconstructed high transverse momentum physics object they are associated with in a chosen order: electrons, photons, hadronically decaying  $\tau$ -leptons, jets and muons. Calorimeter cells not associated with any such objects are also taken into account in the  $E_T^{\text{miss}}$  calculation and they correspond to what is called the “CellOut”  $E_T^{\text{miss}}$  contribution. The full calorimeter  $E_T^{\text{miss}}$  contribution is:

$$E_{x,y}^{\text{miss,calo}} = E_{x,y}^{\text{miss,electrons}} + E_{x,y}^{\text{miss,photons}} + E_{x,y}^{\text{miss},\tau} + E_{x,y}^{\text{miss,jets}} + E_{x,y}^{\text{miss,soft-jets}} + E_{x,y}^{\text{miss,calo } \mu} + E_{x,y}^{\text{miss,CellOut}}, \quad (3.13)$$

where each term is the negative of the sum of calibrated cell energies inside the objects:

$$E_{x,y}^{\text{miss,term}} = - \sum_{i=1}^{N_{\text{cell}}^{\text{term}}} E_i \sin \theta_i \cos \phi_i, \quad (3.14)$$

and  $E_i$ ,  $\theta_i$  and  $\phi_i$  are the energy, polar angle and azimuthal angle of the  $i$ -th cell for all object within  $\eta < 4.5$ . Noise-removal criteria are applied, to reduce the effect of noise in the calorimeter cells [93]. The soft-jets  $E_T^{\text{miss}}$  term is calculated in cells for clusters associated with jets that have a transverse momentum between 7 GeV and 20 GeV and the “CellOut” term includes the energy of cells not associated with any physical reconstructed object. The muon calorimeter term is calculated by matching the muon tracks to the calorimeter. The Muon Spectrometer contribution to the missing transverse energy calculation is estimated from the transverse momentum of muons with tracks within  $|\eta| < 2.7$ , by adding the negative of their transverse momentum components as in:

$$E_{x,y}^{\text{miss},\mu} = - \sum_{\text{muons}} p_{x,y}^{\mu}. \quad (3.15)$$

More information about the  $E_T^{\text{miss}}$  calculation can be found in [93].

The performance of the missing transverse energy calculation was tested in  $Z$ -boson decays into two leptons, in which no missing energy should be detected, so it can be verified that there is no bias [93]. Decays of  $W$ -bosons into a neutrino (which results in missing transverse energy) and a lepton are also used to test the  $E_T^{\text{miss}}$  calculation in [93]. One important measurement of the  $E_T^{\text{miss}}$  performance is its resolution, which follows an approximate stochastic behaviour that can be parametrised as a function of the sum of the transverse energy in the detector:

$$\sigma(E_T^{\text{miss}}) = k \sqrt{\sum E_T}, \quad (3.16)$$

where  $\sigma(E_T^{\text{miss}})$  is the  $E_T^{\text{miss}}$  resolution,  $k$  is a fitted parameter and  $\sum E_T$  is the scalar sum of the transverse energy in the detector. A measurement of the  $E_T^{\text{miss}}$  resolution in  $\sqrt{s} = 7$  TeV ATLAS data and an estimate of the uncertainty

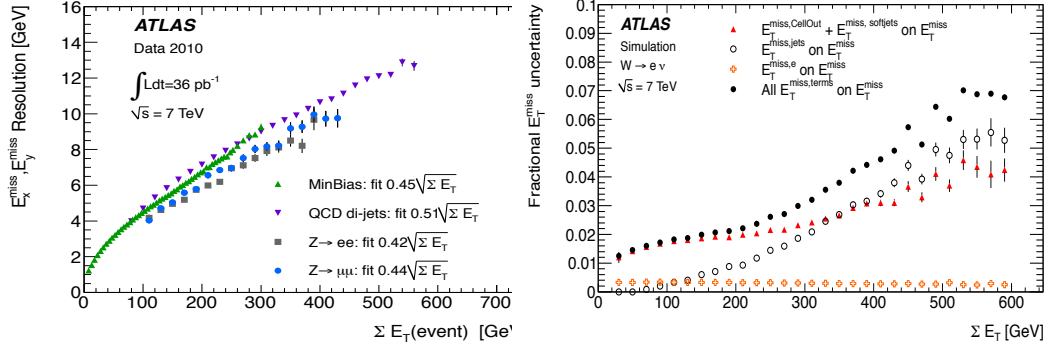


Figure 3.15: Left: resolution of the missing transverse energy measured in 2010 ATLAS  $\sqrt{s} = 7$  TeV data with the respective fits in each channel. Right: uncertainty in the missing transverse energy scale from Monte Carlo simulation of  $W$ -boson decays into an electron and a neutrino. Extracted from [93].

in the  $E_T^{\text{miss}}$  calculation from simulation of  $W$ -bosons to electron and neutrino decays are shown in Figure 3.15.

In the physics analyses that follow, the electrons', muons' and jets' four-momenta are corrected as described in the previous sections of this chapter, to account for discrepancies between data and the simulation. After these corrections, the missing transverse energy is recalculated to obtain a consistent estimate of the  $E_T^{\text{miss}}$ . Uncertainties in the missing transverse energy calculation are also taken into account, by varying the  $E_T^{\text{miss}}$  calculation by the fractional uncertainty in its terms and verifying the resulting impact in the final analyses goals.

### 3.8 Summary

This chapter has summarised a few important aspects of the ATLAS detector, which is used in the analyses that follow. The geometry of the detector and its performance for the identification of each of the relevant physics elements are fundamental to understand how the analysis works.

# Chapter 4

## *b*-jet trigger performance studies

The *b*-jet trigger selects events containing *b*-quarks that hadronise into jets.

Selecting *b*-jet enriched samples is essential for many interesting physics analyses. Searches and measurements involving top quarks rely on a good identification of *b*-jets, since, as has been discussed in Chapter 2, the top quark decays mainly into a *W*-boson and a *b*-quark, due to the large value of the  $|V_{tb}|$  matrix element [7]. Besides top quark-related analyses, SM Higgs boson decays into bottom-antibottom quark pairs is an important signature, since this decay has a large branching ratio [94] ( $\sim 65\%$  for a Higgs mass of 120 GeV, according to [94]).

However, due to the limited trigger bandwidth and storage facilities, one must be very selective on the choice of events to analyse. Trigger chains that select *b*-jets are, therefore, an important element of the ATLAS Trigger system.

### 4.1 *b*-jet trigger chains configuration

The *b*-jet trigger signature is seeded by calorimeter-level seeds, in which a significant transverse energy deposition was measured. In the ATLAS Trigger, this is implemented by the Level 1 jet trigger setup. The Level 2 and Event Filter *b*-jet selection use the Inner Detector tracks that match the equivalent calorimeter region to apply a *b*-tagging algorithm. It is important, therefore, to notice that, although the algorithm is seeded by an energy deposition in the calorimeters, the *b*-tagging algorithm uses mainly tracks.

The part of the *b*-jet High-Level Trigger algorithm exclusively dedicated to *b*-tagging is identical in the Level 2 and in the Event Filter, with the only difference being the usage of the Level 2 tracking algorithm or the Event Filter tracking algorithm, with the calibration of the *b*-tagging algorithm adapted to

the Level 2 or Event Filter tracking performance. The basic structure involves a Feature Extraction algorithm that selects the trigger tracks and calculates a *b*-tagging weight (as described in the next section) using the selected tracks; and a Hypothesis testing algorithm that demands that the *b*-tagging weight is above a certain threshold. The Hypothesis algorithm is configured in three different ways for three different types of trigger chains, which differ in their requirements for the *b*-tagging weight. The “loose” chains apply a requirement on the *b*-tagging weight, so that the efficiency of *b*-jet selection is  $\sim 40\%$ ; the “medium” chains have an efficiency of *b*-jet selection of  $\sim 50\%$ ; while the “tight” chains, have a  $\sim 60\%$  efficiency selection. There are many chains for each of these types, requiring different jet transverse momentum selections, and other chains also make a requirement on lepton triggers.

A set of calibration chains are also available and they are frequently referred to as  $\mu$ -jet trigger chains. These chains use a simpler mechanism to select *b*-jet-enriched samples: they rely on the semileptonic decays of the *B*-hadron into muons. They work by demanding that both a muon and a jet trigger chain passes, and that the muon and jet directions are the same region of the detector. The latter requirement is implemented by demanding that  $\Delta R(\mu, \text{jet}) < 0.4$ . Some of the  $\mu$ -jet trigger chains also restrict the difference in the *z*-coordinate of the jet and muon impact parameters. To differentiate the  $\mu$ -jet calibration trigger chains from the ones used in physics analysis, the former are called “calibration triggers” in this chapter, while the latter are called “physics triggers”.

The *b*-jet physics trigger, chains use selected tracks which are required to have a minimum transverse momentum of 1 GeV, a maximum  $d_0$  of 1 mm, and a  $z_0$  relative to the primary vertex not exceeding 2 mm. The track fit quality must also be good and match the calorimeter-level seed. The *z* position of the primary vertex is calculated by histogramming all selected tracks and using a sliding window algorithm to select the local maximum.

The methods used to detect a *b*-jet rely on the fact that the *B*-hadron resulting from the *b*-jet hadronisation has a relatively long lifetime and it would travel in the detector until it decays in a secondary vertex. Figure 4.1 shows a simplified drawing representing the *B*-hadron decay, in which the *B*-hadron decays in a displaced secondary vertex. This, together with the large mass of the *B*-hadron, generates tracks with high impact parameter (relative to the primary vertex), compared to light-jets. The invariant mass of the secondary vertex can also be used as part of the *b*-tagging procedure. The multiplicity of

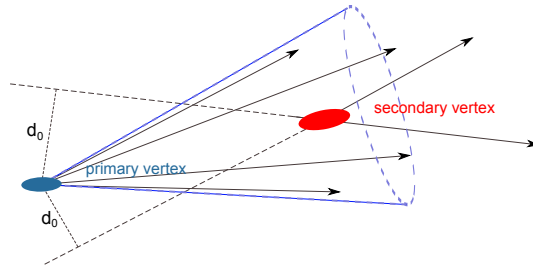


Figure 4.1: Simplified representation of the  $B$ -hadron decay in a secondary vertex. Extracted from [77].

tracks coming from the  $B$ -hadron is also used as part of the  $b$ -tagging procedure. The next section describes the algorithm used to calculate the  $b$ -tagging weights in the  $b$ -jet physics trigger chains in more details.

## 4.2 $b$ -tagging algorithms

Many  $b$ -tagging procedures can be used and there has been a change in the method implemented in the ATLAS  $b$ -jet trigger. The JetProb algorithm was implemented in both Level 2 and Event Filter for the 2011 data collection. The 2012 data collection used a combined tagger, which was calibrated based on the studies that are described in the next sections. The JetProb procedure relies on selecting good quality tracks that match the calorimeter region and only uses the significance of their transverse impact parameter  $d_0$ , defined by  $S(d_0) = \pm \frac{d_0}{\sigma(d_0)}$ . The sign of the impact parameter significances (both  $S(z_0)$  and  $S(d_0)$ ) is distributed according to whether the tracks cross the jet cone axis in front of the primary vertex (positive) or behind it (negative). The algorithm calculates the probability that each track comes from the primary vertex using a fitted resolution function  $\mathcal{R}(x)$ , which is estimated as [77]:

$$\mathcal{P}_{\text{track}} = \int_{-\infty}^{-|S(d_0)|} \mathcal{R}(x) dx. \quad (4.1)$$

This probability is combined into a single probability that all tracks come from the primary vertex, which is expected to be very close to zero for  $b$ -jets and uniformly distributed between zero and one for light-jets. This method is quite robust, since the  $\mathcal{R}(x)$  distribution is fitted from data using only light-jets (assuming the  $b$ -jet contribution is much smaller than the light-jets one), that is, it does not rely on Monte Carlo simulation.

An attempt has been made to switch to more efficient algorithms in 2012, but other algorithms rely on a good agreement between data and simulation on

many variables, including the transverse impact parameter significance,  $S(d_0)$ , the longitudinal impact parameter significance,  $S(z_0) = \frac{z_0 - z_{pv}}{\sigma(z_0)}$ , and estimates of the kinematics of the tracks at the secondary vertex.

The impact parameter-related algorithms take advantage of the *b*-quark decay topology. It is expected that the *b*-quark should hadronise into a *B*-hadron in a secondary vertex. The jet origin would be displaced, due to the *b*-jet lifetime before decaying into the *B*-hadron and the tracks' perigee to the primary vertex, measured by their impact parameters  $z_0$  and  $d_0$  and their uncertainties  $\sigma(z_0)$  and  $\sigma(d_0)$ , would show this displacement relative to the jet cone axis as an asymmetry in the signed impact parameter significances. The IP3D algorithm builds a simulation-level likelihood of the  $S(d_0)$  and  $S(z_0)$  of tracks in *b*-jets and light-jets. A likelihood-ratio is calculated in real data for all tracks combined and a cut is applied on it, to test the hypothesis that the jet indeed contains a *b*-jet. Given the likelihoods  $P_{b,d_0}(S(d_0))$  and  $P_{b,z_0}(S(z_0))$  that represent the probability of a reconstructed *b*-jet's track to have a transverse significance  $S(d_0)$  and a longitudinal significance  $S(z_0)$ , with similar definitions for the light-jet's tracks,  $P_{l,d_0}(S(d_0))$  and  $P_{l,z_0}(S(z_0))$ , the IP3D weight  $w_{\text{IP3D}}$  is defined as:

$$w'_{\text{IP3D}} \triangleq \prod_{i=1}^{\text{N. of tracks}} \frac{P_{b,d_0}(S(d_0)_i) P_{b,z_0}(S(z_0)_i)}{P_{l,d_0}(S(d_0)_i) P_{l,z_0}(S(z_0)_i)} \quad (4.2)$$

$$w_{\text{IP3D}} \triangleq \log_{10} \left( w'_{\text{IP3D}} \right) \quad (4.3)$$

The secondary vertex algorithms reconstruct the invariant mass of the four-momentum sum of all tracks from the secondary vertex, the energy fraction of tracks coming from the secondary vertex relative to all tracks in the jet, and the track multiplicity for tracks matched to the secondary vertex. These variables are also used in simulation to calculate a likelihood function in signal and background, so that a likelihood-ratio can be estimated and a cut is applied in real data. The algorithm used in the 2012 data taking in ATLAS used a combination of the IP3D and the secondary vertex algorithms described, multiplying them into a single combined weight, on which a selection cut can be applied to verify whether the jet is a *b*-jet or not. The combined weight is defined as:



$$\begin{aligned}
w'_{\text{COMB}} &\triangleq w'_E w'_N w'_M w'_{\text{IP3D}} \\
w_{\text{COMB}} &\triangleq -\log_{10} \left( 1 - \frac{w'_{\text{COMB}}}{1 + w'_{\text{COMB}}} \right),
\end{aligned} \tag{4.4}$$

where  $w'_E$  is the likelihood ratio for the calculated secondary vertex energy,  $w'_N$  is the likelihood ratio for the calculated number of tracks matched to the secondary vertex and  $w'_M$  is the likelihood ratio for the calculated invariant mass of the secondary vertex.

The next section shows the data to simulation agreement for some important variables, which are used to establish that there is indeed a good agreement between data and simulation when switching to the new algorithm in the 2012 data taking. The data to simulation comparison of the combined tagger used in the 2012 data taking is shown in the final section.

### 4.3 Data to Monte Carlo simulation comparison of with $\sqrt{s} = 7$ TeV data

A data to simulation comparison was prepared to check the reconstruction quality of the Event Filter tracks. The Event Filter significance of  $d_0$ , the tracks' multiplicity and the tracks' transverse momentum is shown in Figures 4.2, 4.3, 4.4. These plots show also the individual contribution of b-jets, c-jets and light-jets in simulation. To determine the jet flavour in the simulated events, the Event Filter jets'  $\eta$  and  $\phi$  coordinates were used to match the trigger jets to the Anti- $k_t$   $R = 0.4$  jets calculated by the Offline ATLAS software, which has access to the full detector information, with all jet energy corrections applied. The standard Anti- $k_t$  ATLAS jets were matched to  $B$ -hadrons or  $C$ -hadrons, requiring a  $\Delta R < 0.3$  between them, to verify whether the jet is a  $b$ -jet, a  $c$ -jet, or a light-jet. In these figures, the simulation has been normalised to the total data yield.

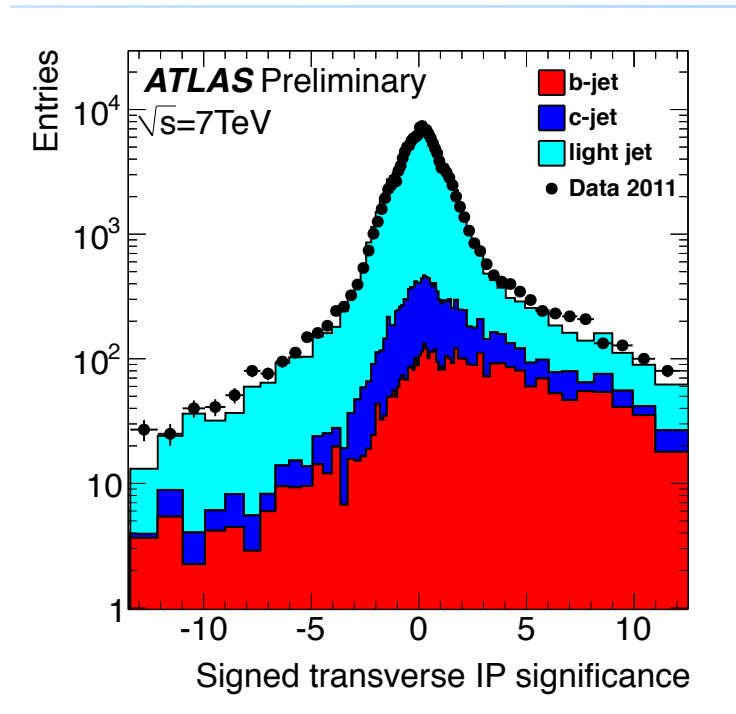


Figure 4.2: The signed  $S(d_0)$  for the selected tracks at the Event Filter, using 2011 ATLAS data.

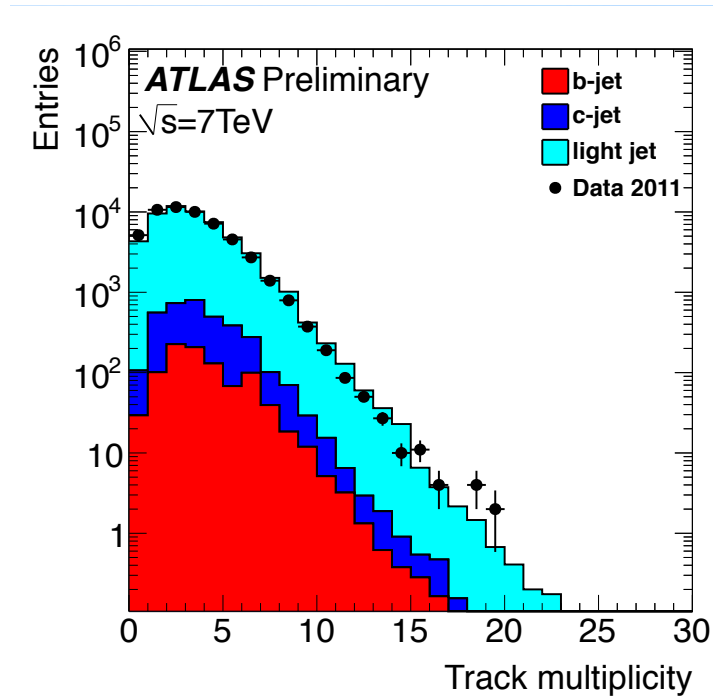


Figure 4.3: The track multiplicity for selected tracks at the Event Filter, using 2011 ATLAS data.

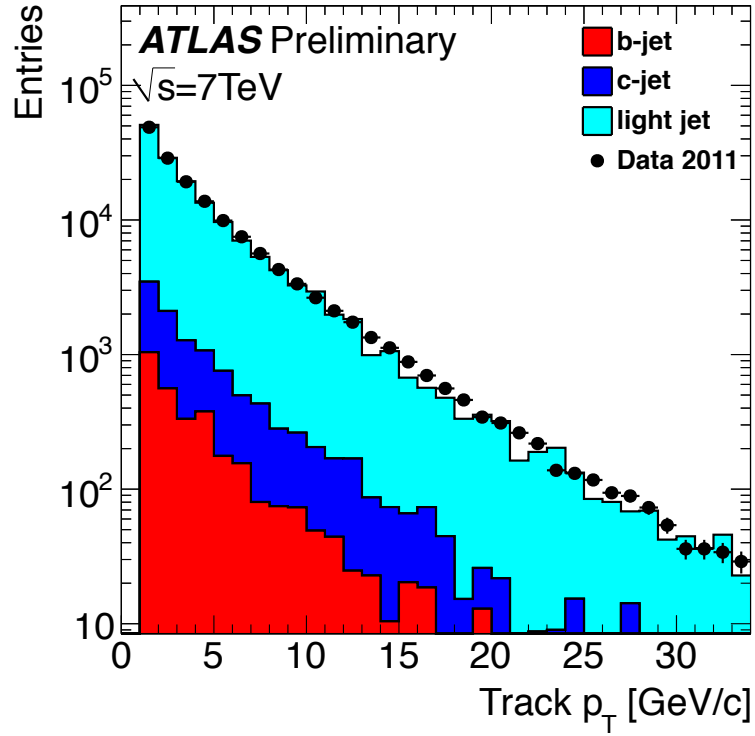


Figure 4.4: The tracks’ transverse momenta for selected tracks at the Event Filter, using 2011 ATLAS data.

#### 4.4 Data to simulation comparison of the $b$ -jet combined “physics” trigger using $\sqrt{s} = 8$ TeV, 2012 data

A similar data to simulation comparison has been done with 2012 ATLAS data, but analysing the final  $b$ -tagging weight. In this comparison, to avoid biasing the comparison in the tagger weight distribution, the trigger selection used applies no  $b$ -tagging cut on the  $b$ -jet candidate in the “physics trigger”<sup>1</sup>, but only a cut on the transverse momentum of the jet candidate.

The transverse momentum cut is applied on all  $b$ -jet physics selection chains, therefore this restriction presents no bias in the result, compared to other  $b$ -jet physics trigger chains. Due to the high rate of events when the  $b$ -tagging cut is removed, this trigger is prescaled on data, which means that there will be a loss of events.

The combined tagger weight, on which the  $b$ -tagging cut is applied was plot, as well as the likelihood-ratio IP3D tagger, the number of tracks at the

<sup>1</sup>In the ATLAS jargon, the trigger EF\_b55\_NoCut\_j55\_a4tchad was used.

Secondary Vertex, the energy fraction at the Secondary Vertex and the invariant mass of the Secondary Vertex, which are the relevant components for the combined tagger. The 2012 ATLAS  $\sqrt{s} = 8$  TeV data runs 213359 up to 213640 were used in this analysis, representing an integrated luminosity of  $5.1\text{nb}^{-1}$ . To understand the flavour fractions in these plots, the true flavour of the simulation-level jets can be investigated. The HLT jets were geometrically matched to the Offline jets reconstructed with Anti-kt  $R = 0.4$  using a  $\Delta R < 0.4$  criterion, considering only  $|\eta| < 2.5$  and  $p_T > 40$  GeV offline jets. The results are shown in Figures 4.5, 4.6, 4.7, 4.8, 4.9. The Offline jets are matched to  $b$ - or  $c$ -hadrons in simulation to classify them as  $b$ -jets,  $c$ -jets or light-jets. This allows one to see the simulation-level flavour fraction in the figures. It can be seen that, for high values of the combined tagger weight, the  $b$ -jet component fraction is increased, as it is expected. The modelling of the variables used in the combined tagger also seems to be consistently done in simulation, when compared to data.

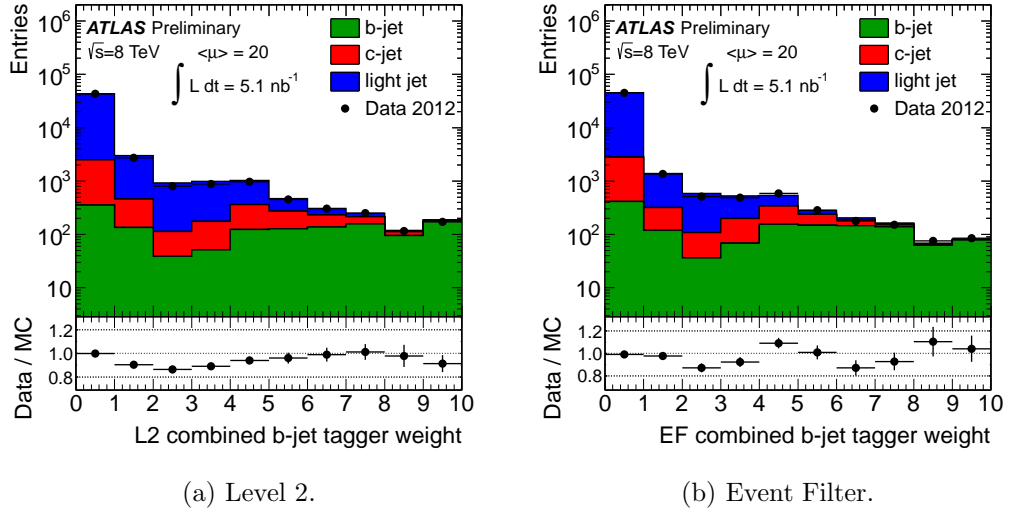


Figure 4.5: Combined tagger weight for physics trigger using the impact parameter significance and the secondary vertex likelihood-based taggers, calculated from Level 2 and Event Filter tracks in low  $p_T$  jets identified by the Level 1.

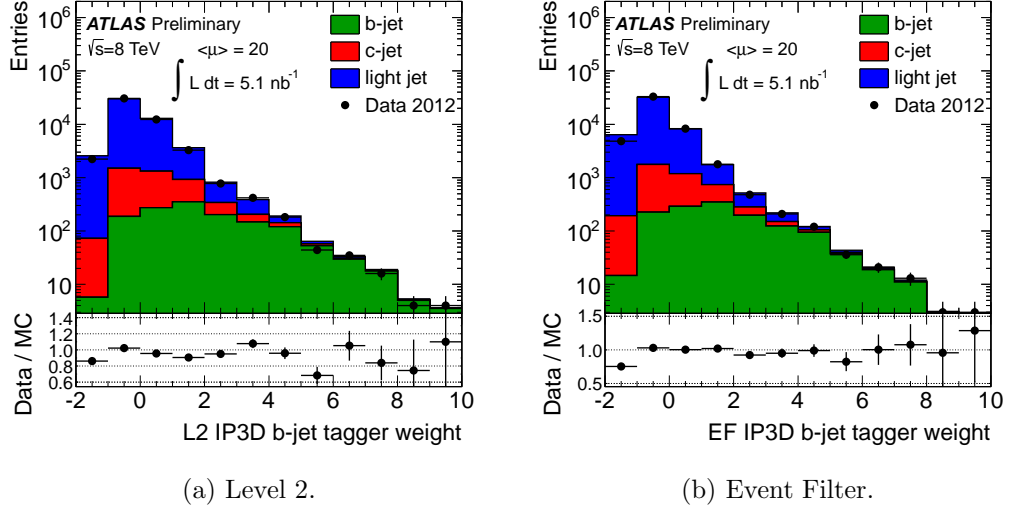


Figure 4.6: IP3D tagger for physics trigger using the transverse and longitudinal impact parameter significances, calculated using Level 2 and Event Filter tracks in low  $p_T$  jets identified by the Level 1.

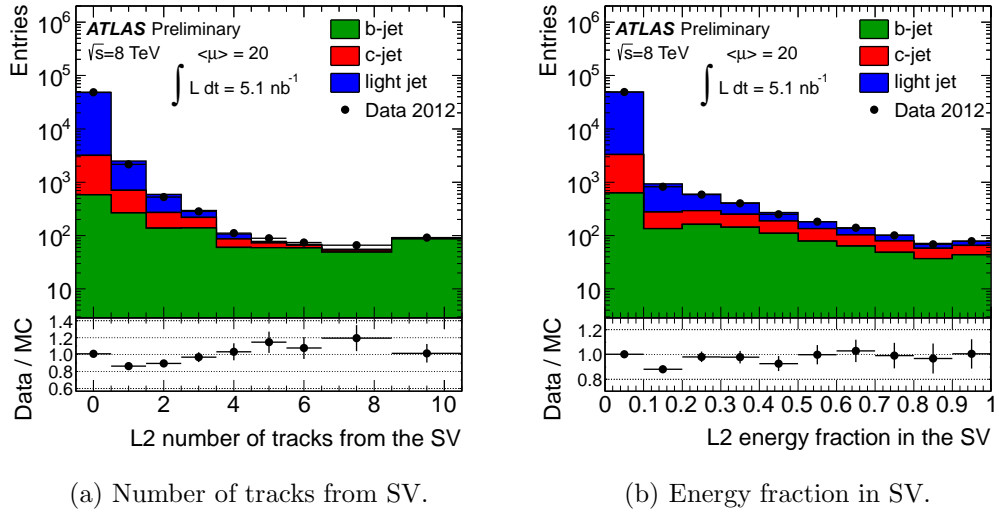


Figure 4.7: Data to simulation comparison for the physics trigger with flavour association for the Level 2 SV variables.

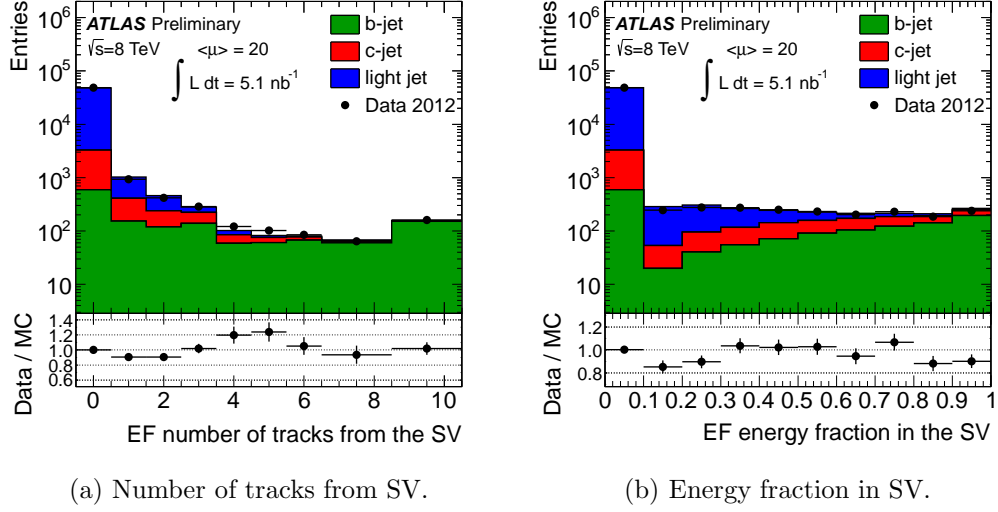


Figure 4.8: Data to simulation comparison for the physics trigger with flavour association for the Event Filter SV variables.

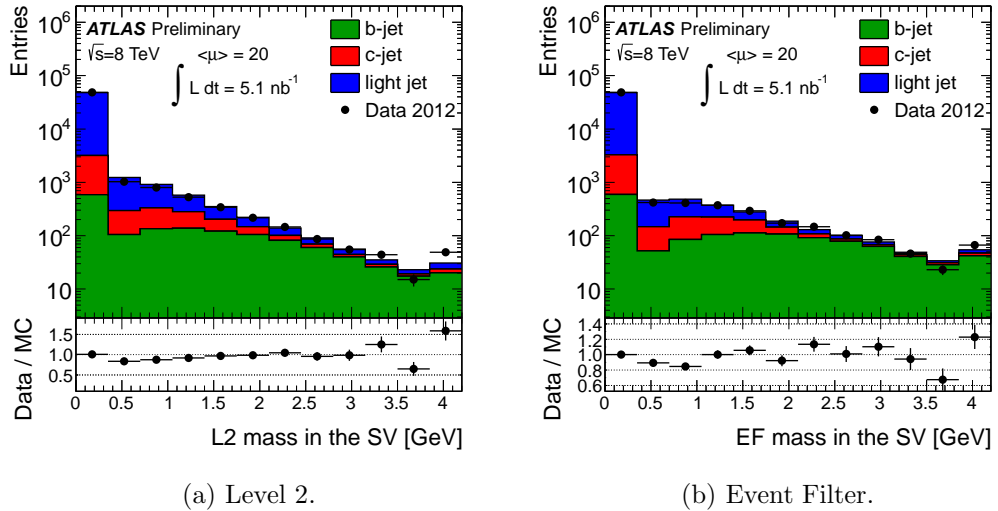


Figure 4.9: Data to simulation comparison for the physics trigger with flavour association for the mass of the SV.

## 4.5 Data to simulation comparison in heavy flavour enriched sample with ATLAS 2012 $\sqrt{s} = 8$ TeV data

A set of calibration triggers were also developed to get a sample of events enriched with heavy flavour content, using an orthogonal selection to the main

$b$ -jet triggers. The idea relies on the semileptonic decay of the  $b$ -hadrons. This allows the selection of a heavy flavour enriched sample, by selecting jets which have muons nearby with  $\Delta R < 0.4$ .

For the analysis performed, a similar set of cuts was used, compared to the “physics” triggers analysis in the previous section, but it was required that a muon and a jet trigger passed and that they are in the same  $\eta \times \phi$  region of the detector <sup>2</sup>. All other selection requirements, were done in accordance to the cuts described in Section 4.4. The data runs used in this analysis range from 213359 up to 213640, representing an integrated luminosity <sup>3</sup> of  $1.4\text{pb}^{-1}$ .

Figures 4.10 and 4.11 show the Level 2 and Event Filter data to simulation comparison using the calibration triggers for the combined and IP3D taggers respectively. It can be noticed that the combined tagger weight has generally a larger fraction of events at higher values, compared to the result in Section 4.4, which means that this analysis is more sensitive to the heavy flavour content on the data sample. The track multiplicity and energy fraction data to simulation comparison are also shown in Figures 4.12 and 4.13 for the Level 2 and Event Filter. The estimated mass of the secondary vertex is shown in Figure 4.14. These studies show that the modelling of the  $b$ -jet component in simulation is consistent with data and that the  $b$ -jet trigger selection enhances the  $b$ -jet component for a cut on the combined tagger weight.

---

<sup>2</sup>In the ATLAS jargon, the trigger `EF_mu4T_j55_a4tchad_matched` was required to pass.

<sup>3</sup>Note that the calibration trigger used here and the physics trigger without a  $b$ -tagging selection used in the previous section are both prescaled by different amounts.

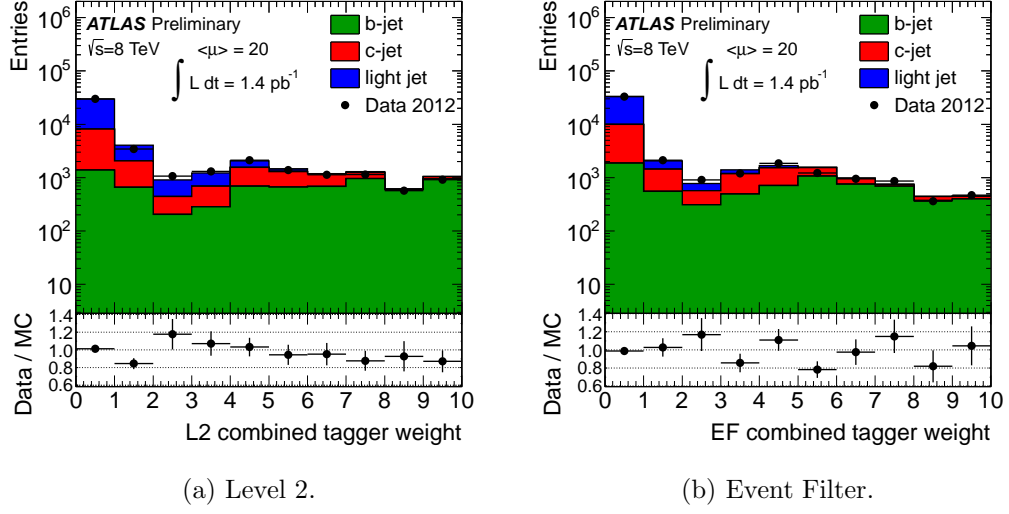


Figure 4.10: Combined tagger weight using the impact parameter significance and the secondary vertex likelihood-based taggers, calculated from Level 2 and Event Filter tracks in low  $p_T$  jets identified by the Level 1.

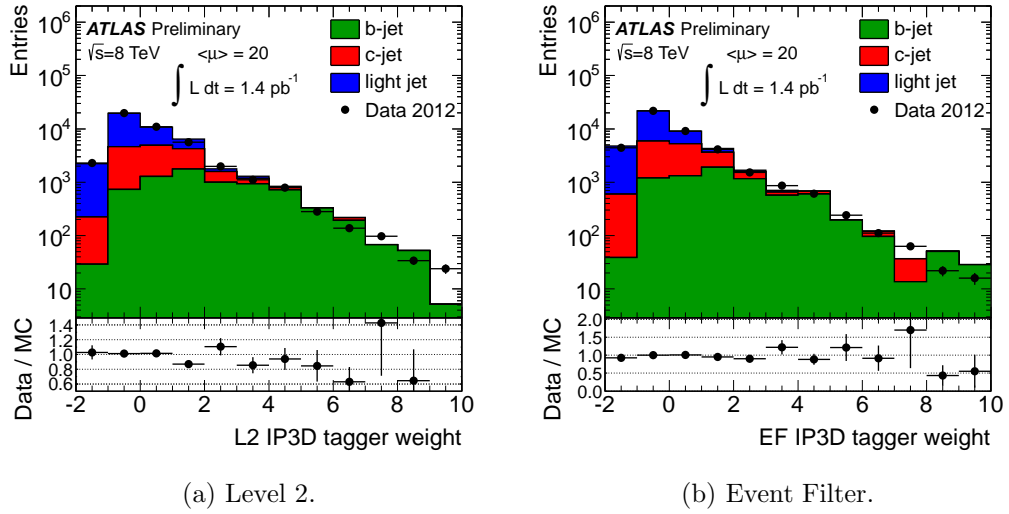


Figure 4.11: Data to simulation comparison for the calibration trigger with flavour association for the IP3D tagger.



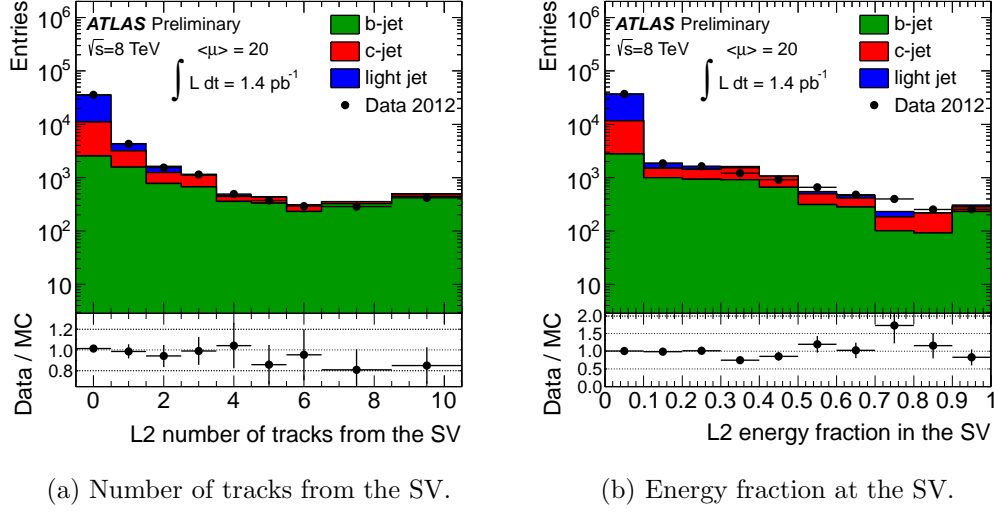


Figure 4.12: Data to simulation comparison for the calibration trigger with flavour association for the Level 2 SV variables.

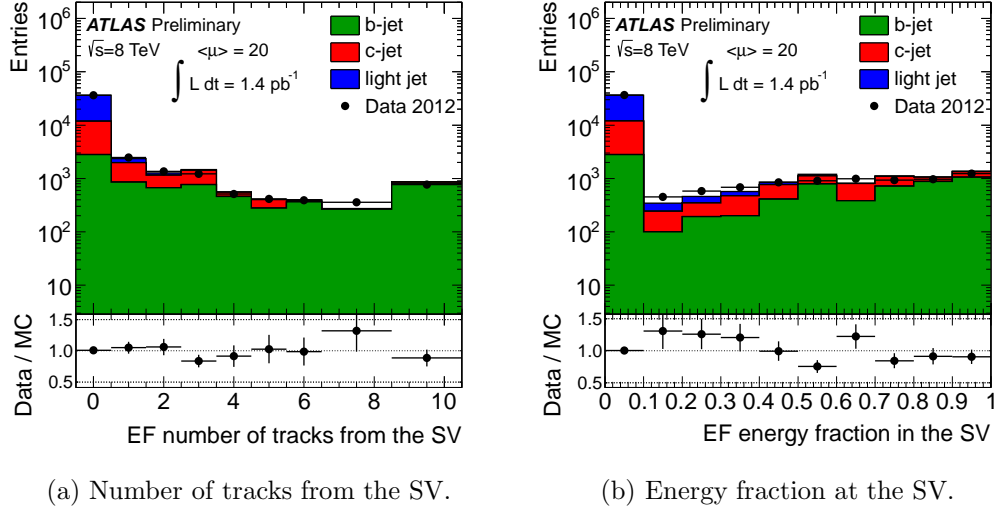


Figure 4.13: Data to simulation comparison for the calibration trigger with flavour association for the Event Filter SV variables.

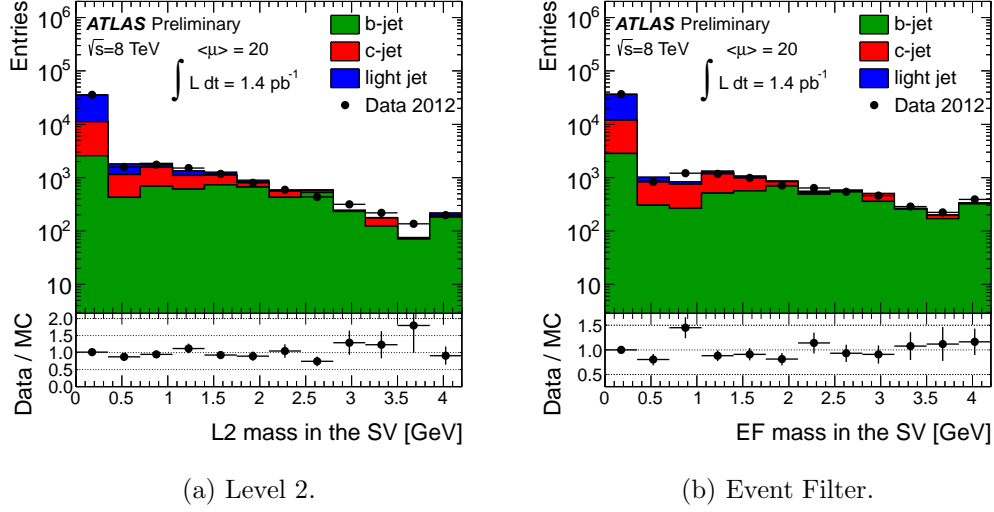


Figure 4.14: Data to simulation comparison for the calibration trigger with flavour association for the mass of the SV.

## 4.6 Summary

The  $b$ -jet trigger allows one to enrich physics analysis samples with heavy flavour, which is very important for many analyses, including top quark and Higgs boson related studies. To perform reliable studies on the ATLAS  $b$ -jet trigger, it is important to test whether the simulation can describe real data well. Furthermore, it is important to test whether the trigger selects heavy flavour events. This is essential for performance studies on the  $b$ -jet trigger. In these studies, it is possible to verify that there is good agreement between data and simulation in the ATLAS data collected in 2011, which allowed the  $b$ -jet trigger to be calibrated to use more complex likelihood-ratio-based taggers.

The data collected in ATLAS in 2012 uses a combined tagger, based on the tracks' impact parameter and the secondary vertex observables. It has been shown that the combined tagger in the physics trigger enhances the  $b$ -jet component for higher combined weight values. The agreement between data and simulation for the heavy flavour enhanced calibration triggers is also very good, showing that the  $b$ -jet component is also correctly modelled.

# Part III

## Physics Analyses

## Chapter 5

# Top-antitop differential cross section measurement as a function of the jet multiplicity in the final state

This study focuses on the semileptonic decays of the top-antitop ( $t\bar{t}$ ) system and aims at measuring its cross section as a function of the number of jets produced in the final state.

The motivation behind this study is discussed in the next section, followed by details about the samples used for the analysis signal and background estimate. The unfolding method used, which corrects for detector effects is presented with a discussion of the systematic effect that it produces. The analysis has been made public as a conference note by ATLAS [95].

### 5.1 Motivation

As mentioned in Chapter 2, the top quark has a very large mass compared to the other quarks and, therefore, it is expected that it has a very large coupling to the Standard Model Higgs boson, which is central to the Electroweak Symmetry Breaking mechanism. Furthermore, the top quark has a very short lifetime and decays before hadronising. In its decay, the top quark is expected to radiate gluons, generating extra jets in the final state. This extra radiation would appear as higher order diagrams, in addition to the lowest order decay and can be predicted using Quantum Chromodynamics. A measurement of the production of this extra radiation is important to test how well the Standard

Model describes the production of both the top quark and the associated QCD radiation.

Furthermore, the top-antitop production is a main background in many physics analyses, such as the  $t\bar{t}H$  search and in searches of supersymmetric models. A good understanding of  $t\bar{t}$  production is essential to have a good modelling of other analyses' backgrounds.

In the current analysis, it is proposed to measure the production cross section of the  $t\bar{t}$  system as one of the top quarks decays leptonically and the other top quark decays hadronically. The measurement of the production cross section is done as a function of the number of jets in the final state, which are related to the amount of radiation produced in the  $t\bar{t}$  decay. The analysis is, therefore, an important test of the Standard Model description of the top quark. The main elements of the analysis include the signal selection, which improves the signal-to-background ratio, the background estimation and the unfolding procedure (described in Section 5.9), which corrects for the detector effects.

In the next section an account of the signal modelling through different simulation methods used in the analysis is given. The signal modelling is important to estimate the detector effects which need to be corrected in the unfolding procedure (Section 5.9). The background is simulated or estimated through data-driven methods and it is also discussed in the next sections. The event selection is discussed next, which attempts to mitigate the effect of the background. The unfolding procedure is detailed then, followed by the method used for uncertainty propagation. Finally the resulting data to simulation comparison before and after the unfolding is shown.

## 5.2 Top-antitop signal simulation and background estimates

The signal is simulated as it is understood in the Standard Model to measure the effect of the detector. It is also simulated using different generators and configurations to test the agreement with data after the effect of the detector is corrected for. All sources of background are estimated as well, so that this contribution can be subtracted from data.

As a first step, the signal must be defined in terms of its final state objects, so that the selection can be properly justified. The signal contains an electron, muon or tauon and a neutrino as a result of the  $W$ -boson decay, two  $b$ -quarks

from the top decays and two other quarks from the other  $W$ -boson decay. The  $\tau$  lepton decays either leptonically into an electron or muon and neutrinos, or it decays hadronically, generating jets. If the  $\tau$  lepton decays hadronically, it will generate jets, which we do not intend to select and it will be considered as part of the background.

Therefore, the detector measures an electron or muon, missing transverse energy due to the undetected neutrino and at least four jets, since more jets could be produced as a result of QCD radiation. As there is either an electron or a muon in the final state, this splits the signal into two distinct channels to be taken into consideration, which are referred to henceforth as the “electron channel” and the “muon channel”. In this chapter, for what follows, the word “lepton” will be used to refer only to the final state electron or muon, with the possible  $\tau$  leptonic decays being indirectly included as electron or muon decays.

The  $t\bar{t}$  signal is simulated with different fixed order calculations and matching schemes for the parton shower. The Alpgen v2.13 generator [44] with the CTEQ6L1 [38] PDF set is used as the main reference sample to derive correction factors in the unfolding procedure. This sample was chosen since it predicts the jet multiplicity distribution well at reconstruction-level. A systematic uncertainty associated with the choice of this sample is calculated at a later stage.

The Alpgen reference sample is generated with zero up to four exclusive and five inclusive additional partons produced as extra radiation. Herwig v6.520 [48, 49] is used for the parton showering and the fragmentation and the MLM [52] parton-jet matching scheme is applied<sup>1</sup> to avoid double counting between the matrix element calculation and the parton shower. MC@NLO [56]<sup>2</sup> and Powheg [54]<sup>3</sup> generators are used when comparing the final unfolded result in Section 5.11. MC@NLO is also used to estimate the systematic effect caused by the unfolding procedure. Jimmy [63] is used with each sample produced with Herwig, for the underlying event simulation with the AUET1 tune. An Alpgen v2.14 sample is also produced using Pythia [50] for the parton shower and the CTEQ5L PDF set to test the systematic effect of the parton shower. Furthermore, to test the effect of the ISR/FSR models, Alpgen samples are generated with different renormalisation scales associated with doubling and

---

<sup>1</sup>With parameters ETCLUS 20 GeV, RCLUS 0.7, ETACCLUS 6.0.

<sup>2</sup>Using the CT10 [39] PDF set and interfaced with Herwig for the parton shower.

<sup>3</sup>Using the CTEQ 6.6 [38] PDF set and interfaced with Pythia [50] for the parton shower, using the AUET2B-CTEQ 6L1 tune.

halving the scale  $Q$  at which  $\alpha_S(Q)$  is calculated in the matrix element, while keeping the  $\alpha_S$  configuration for the parton shower and the PDF set fixed. The nominal  $\alpha_S$  value is set to be 0.118 at the  $Z$  boson mass scale. The Alpgen+Pythia nominal and  $\alpha_S$  variation samples are produced using the Pythia Perugia 2011 tune [96]. The top mass was set to be  $m_{t\bar{t}} = 172.5$  GeV and its production cross section was normalised to  $\sigma_{t\bar{t}} = 167_{-18}^{+17}$  pb in all samples used<sup>4</sup>.

Many background events can be selected as they have a similar final state configuration as the signal, either because the final state particles are the same, or because there are experimental effects which lead the selection mechanism to tag the background as the signal, such as misidentification of jets as leptons. The selection procedure in this analysis emphasizes the  $t\bar{t}$  signal over the background, rejecting significantly more background than signal, but it is not enough to completely remove the background. It is necessary to estimate the remaining background in the data after the selection has been made, and subtract them.

The main background to the  $t\bar{t}$  production estimate in the semileptonic channel is  $W$ +jets production, which could be falsely identified as being part of the signal if the  $W$  boson decays leptonically and there are extra jets produced by radiation. The shape of the  $W$ +jets kinematic distributions is estimated from Alpgen v2.13, using the CTEQ6L1 PDF set and interfaced with Herwig for the parton shower, but the normalisation of the distribution and the heavy flavour content is not well predicted from the simulation and a partially data-driven method is used. The  $W$ +jets estimate also includes an estimate of the heavy and light flavour content partially calculated from data, which is important to understand the effect of the  $b$ -tagging cut on the  $W$ +jets distribution. This method is explained in more details in Section 5.4. It is worth noting that, since the samples of  $W + b\bar{b}$ + jets,  $W + c\bar{c}$ + jets,  $W + c$ + jets and  $W$ + light-jets are produced separately, a heavy flavour overlap removal procedure is used to remove the overlap between the heavy flavour content of the samples, using a  $\Delta R$  match between the simulation-level Anti- $k_t$  jets and the reconstruction-level Anti- $k_t$  jets and removing the event if they represent an overlap between the samples.

Due to the high number of jets involved in the analysis, the QCD multi-jet production is very low. Therefore, the QCD multi-jet production simulation cannot be used to generate enough events and other methods are used to

---

<sup>4</sup>Obtained from approximate NNLO QCD [97] calculations at the  $m_{t\bar{t}}$  point used.

estimate it. The main source of misidentification of QCD multi-jet events as  $t\bar{t}$  decay products happens if one of the jets is misidentified as a lepton. This motivated a data-driven estimate of this background's contribution, which is done by estimating how often one of the jets in a QCD enriched data sample is mistaken for a lepton. This method was used in this analysis and details on how it was done are mentioned in Section 5.5.

The  $Z$ +jets sample is simulated using Alpgen v2.13 with the CTEQ 6L1 PDF set, using Herwig for the parton shower. An angular matching is used, as in the  $W$ + jets case to remove the overlap between the  $Z$ +jets and the  $Z + b\bar{b}$ +jets samples. The  $t$ -channel single top quark sample was generated using the AcerMC [43] generator, while the  $Wt$ - and  $s$ -channel predictions were simulated using MC@NLO. Each single top sample was normalised to the approximate NNLO calculation of each respective channel [98–100]<sup>5</sup>. The  $WW$ ,  $ZZ$  and  $WZ$  production are generated using Herwig.

## 5.3 Top-antitop event selection

In what follows, the requirements used to select events are described. The cuts used to select reconstruction-level events are organised in a set of main items: a trigger-related selection, lepton selection, jet selection and missing energy requirements.

### 5.3.1 Trigger and pile up-related selection

A first step in the selection procedure is to use the trigger to select events tagged as containing at least one electron with  $p_T > 20$  GeV or  $p_T > 22$  GeV depending on the relevant data taking period, or at least one muon with  $p_T > 18$  GeV, according to the selection channel<sup>6</sup>. A second step demands that at least four tracks were used to reconstruct the position of the primary vertex, which works as a quality cut on the event reconstruction and it reduces the effect of multiple proton-proton interactions.

---

<sup>5</sup> $\sigma_{t,t\text{-channel}} = 64.5^{+2.6}_{-1.7}$  pb,  $\sigma_{t,s\text{-channel}} = 4.6^{+0.2}_{-0.2}$  pb,  $\sigma_{t,Wt\text{-channel}} = 15.7^{1.2}_{-1.2}$  pb.

<sup>6</sup>EF\_e20\_medium for periods B-J, EF\_e22\_loose for period K and EF\_e22vh\_medium1 or EF\_e45\_medium1 for periods L to M, in the electron channel; EF\_mu18 for periods B to I and EF\_mu18\_medium for periods J to M, in the muon channel.



### 5.3.2 Lepton selection

As a next step, it is important to demand that one and only one well identified and isolated electron or muon was detected in the electron or muon channel, respectively. A few quality cuts must be required for the lepton, to remove misreconstructed leptons or non-prompt leptons, which come from the decay products of other particles and not from the top quark's  $W$  boson decay. The electrons are required to have a transverse momentum of at least 25 GeV and it must have a well reconstructed track that matched a calorimeter energy deposition region. Electrons are also required to be in the pseudorapidity range  $|\eta| \in [0, 1.37) \cup (1.52, 2.47]$  to exclude the “crack” region of the calorimeters, in which the energy of the electron is not well reconstructed. For muons, the Inner Detector track is required to match the Muon Spectrometer track and it must be in the pseudorapidity range of  $|\eta| < 2.5$ . Furthermore, the sum of energy in the calorimeters in a region of  $\Delta R < 0.2$  around the muon, excluding its own energy is required to be less than 4 GeV and the sum of the track momenta in a region of  $\Delta R < 0.3$  (excluding the muon's track momentum) is also required to be less than 4 GeV. For the electron the same energy sum and track momenta sum is calculated, but the cut is adjusted for each particle, so that the cut has a flat 90% efficiency in the electron's transverse momentum <sup>7</sup>.

An important quality cut to be taken into account is to reject electrons that overlap with jets if  $0.2 < \Delta R(e, \text{jet}) < 0.4$  and remove jets that satisfy  $\Delta R(e, \text{jet}) < 0.2$ . The latter removal step attempts to reduce the effect of leptons which were misidentified as jets, while the former electron removal step is necessary because the electron identification and reconstruction corrections to be applied at a later stage assume well separated electrons and jets. A similar overlap removal quality cut is applied to muons, which are required to be far away from jets, according to the criteria  $\Delta R(\mu, \text{jet}) > 0.4$ , removing non-prompt muons that might be produced from the semileptonic  $b$ -quark decay and that would be close to the  $b$ -jet.

It is also important to reject the event if a lepton of a different type is found in the event, that is, in the electron channel, no good quality muons should exist and in the muon channel, no good quality electrons should exist, which removes the effect of the dilepton top-antitop decays and  $Z$  boson decays into two leptons. To increase the purity of the selected sample, if there is at least

---

<sup>7</sup>In the ATLAS jargon, the variables calculated are called **Etcone20** and **Ptcone30**. In the muon channel, it is demanded that **Etcone20** < 4 GeV and **Ptcone30** < 4 GeV, while for the electron channel the cuts are adjusted so that an efficiency of 90% is kept as a function of transverse momentum.

one lepton of the other type with a transverse momentum greater than 15 GeV, the event is rejected <sup>8</sup>. The selected lepton is also required to match the lepton selected by the trigger system in the corresponding electron or muon trigger selection <sup>9</sup>.

### 5.3.3 Jet selection

Although the total number of jets from the  $t\bar{t}$  decay channel of interest includes at least four jets, some of the jets might not be detected within the fiducial volume of the detector. The unfolding procedure could benefit from an estimate of these jets being lost, by including the events with three jets as well, so that migrations of events from the four-jets category to the three-jets category could be corrected. Therefore, the event is required to contain at least 3 jets within  $|\eta| < 2.5$  and a jet vertex fraction (see Section 3.5 for the definition of the jet vertex fraction) greater than 0.75 to reduce the pile up contribution. The transverse momentum cut on each jet is chosen to be 25 GeV, 40 GeV, 60 GeV and 80 GeV in four versions of this analysis, that is, the analysis is repeated four times, each one demanding that all selected jets have a specific minimum energy. This procedure shows the effect of jets at different transverse momenta ranges in the jet multiplicity distribution. The particular choice of these values is arbitrary, though.

### 5.3.4 Missing energy requirements

To reject backgrounds without neutrinos in the final state, such as the fully hadronic decay of the top-antitop pair and the QCD multi-jets background, the total missing transverse energy in the event is required to be greater than 30 GeV.

The transverse mass between the lepton and the missing transverse energy,  $E_T^{\text{miss}}$ , is defined as:

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \alpha)}, \quad (5.1)$$

where  $p_T^\ell$  is the lepton's transverse momentum,  $E_T^{\text{miss}}$  is the missing transverse

---

<sup>8</sup>Note that this is smaller than the 25 GeV requirement for the selected lepton.

<sup>9</sup>It is required to match one of the selected trigger electrons (for the electron channel) or muons (for the muon channel), since another object might also have been accepted by the trigger system as an electron or muon. This would happen because the trigger system might misidentify one or more of the objects. No demand is made on the multiplicity of the selected trigger electrons or muons.

energy, and  $\alpha$  is the azimuthal angle between  $p_T^\ell$  and  $E_T^{\text{miss}}$ . This variable has an end-point at the mother particle's mass, which, for a top-antitop decay, should be the  $W$ -boson mass. Taking advantage of this, a cut is applied on  $m_T$ , which is required to be greater than 35 GeV. Finally, at least one of the jets, with  $p_T > 25$  GeV is required to be  $b$ -tagged. The  $b$ -tagging criteria in the analysis was chosen based on the 70% efficiency operating point of the Neural Network based tagger MV1 (see Section 3.6 for more information on  $b$ -tagging). One further  $b$ -jet could have been required, since in principle  $t\bar{t}$  events should contain at least two  $b$ -jets, but due to the low efficiency of the  $b$ -tagging algorithm<sup>10</sup>, making extra demands on the  $b$ -tagging criteria would increase the statistical uncertainty significantly.

## 5.4 Data-driven $W$ +jets background estimate

As mentioned previously, the  $W$ +jets production rate and flavour fractions are not well described in simulation and a partially data-driven method is used in this work. This method was applied in this and other top quark related analyses and the estimate of the factors mentioned in this section were performed by other ATLAS working groups and not the author of this document himself. The method is briefly introduced for clarity and completeness only.

The data-driven  $W$ +jets estimate is performed in three stages. In the first step, the flavour fraction of the  $W$ +jets background is constrained before the  $b$ -tagging requirement. A control region is defined with the same selection cuts, but constraining the jet multiplicity in the final state to be one or two jets, instead of at least three jets. The control region described is enriched in the  $W$ +jets background and the small contribution from other backgrounds and  $t\bar{t}$  is subtracted. The backgrounds and  $t\bar{t}$  contributions' are estimated using Monte Carlo simulation in the Control Region. This results in a final event count for the 1- and 2-jet channels, before and after the  $b$ -tagging selection, which will be referred to as  $W_{i,\text{pre-tag}}^{\text{data}}$  and  $W_{i,\text{tagged}}^{\text{data}}$ , for  $i = 1, 2$ .

Equations can be written for  $W_{1,\text{pre-tag}}^{\text{data}}$ ,  $W_{1,\text{tagged}}^{\text{data}}$ ,  $W_{2,\text{pre-tag}}^{\text{data}}$  and  $W_{2,\text{tagged}}^{\text{data}}$  as a function of eight independent flavour fractions as unknowns, which represent the fractions of  $W + b\bar{b} + \text{jets}$ ,  $W + c\bar{c} + \text{jets}$ ,  $W + c + \text{jets}$  and  $W + \text{light-jets}$  in the 1- and 2-jet bins before  $b$ -tagging. The  $b$ -tagging probabilities estimated in simulation are used to express the tagged quantities as a function of the

---

<sup>10</sup>Notice that increasing the efficiency point would also increase the false-identification rate, reducing the purity of the selected sample.

untagged flavour fractions in the relevant equations. The ratio between the 1- and 2-jet bin event counts are calculated from simulation to relate the values in the set of the equations. Furthermore, the ratio of number of events coming from  $W + c\bar{c} + \text{jets}$  and  $W + b\bar{b} + \text{jets}$  is estimated in simulation to reduce even more the number of degrees of freedom in the equations. With these constraints, the three equations remain with only three degrees of freedom, which can be extracted, by solving the linear system. This procedure results in an estimate for the flavour fractions before  $b$ -tagging for the  $W + b\bar{b} + \text{jets}$  (which are related to the  $W + c\bar{c} + \text{jets}$  by the ratio estimated in simulation), the  $W + c + \text{jets}$ , and the  $W + \text{light-jets}$ . These fractions are applied to the relevant simulation data samples to improve the flavour fraction components from this data-driven measurement, but no normalisation change is made at this stage.

The second step in this estimate is to use a data-driven method to calculate the normalisation of the  $W$ +jets background after the selection but without the  $b$ -tagging requirement. It can be done by noting that the production rate of  $W^+ + \text{jets}$  is bigger than that of  $W^- + \text{jets}$ , since there are more up valence quarks in the protons than down valence quarks and that the ratio between the production rates of  $W^+ + \text{jets}$  and  $W^- + \text{jets}$ ,  $r_{MC}$ , is more precisely calculated from simulation than the rates themselves. The estimate is done by counting the number of data events in the signal region which produce positively charged leptons and negatively charged leptons, that is, the events that come respectively, from the  $W^+$  and  $W^-$  decays. In addition, other processes involved in this measurement produce, to good approximation, equal number of positively and negatively charged leptons and, since we are only interested in the asymmetry between these rates, they would not interfere in this estimate. The remaining background is subtracted. The number of  $W$ +jets events before the  $b$ -tagging requirement can, then, be calculated using Equation 5.2, in which  $r_{MC}$  is the ratio between  $W^+$  and  $W^-$  events produced in simulation,  $D_{W^+}$  and  $D_{W^-}$  are the amount of detected  $W^+$  and  $W^-$  in data and  $W_{\text{pre-tag}}$  represents the total number of  $W$ +jets events produced (before the  $b$ -tagging requirement).

$$W_{\text{pre-tag}} = N_{W^+} + N_{W^-} = \left( \frac{r_{MC} + 1}{r_{MC} - 1} \right) (D_{W^+} - D_{W^-}) \quad (5.2)$$

Each  $W$ +jets flavour component sample is simulated after the selection including the  $b$ -tagging requirement, with each data sample weighted by the estimated flavour fraction. The final step estimates the effect of the  $b$ -tagging

requirement in simulation and applies it to the previously calculated normalisation [95, 101].

## 5.5 Data-driven QCD multi-jets background estimate

The calculation of the data-driven QCD multi-jets estimate was not done by the author in this analysis<sup>11</sup>, but the procedure is mentioned here for completeness, as it was for the case for the W+jets estimate in the previous section. The QCD multi-jets background is estimated by a method called Matrix Method, which associates weights to data events with looser lepton identification requirements to generate distributions for this background.

Events for this QCD estimate are categorised as containing a “loose” lepton, which only satisfy a looser lepton identification criteria, with no isolation requirement, or they are categorised as containing a “tight” lepton, which also satisfies the standard selection lepton identification criteria. The real data estimate and the Monte Carlo simulation estimates in the analysis contains only events that pass the event selection using the “tight” lepton criteria. The QCD estimate described below contains real data events that pass the selection using the “tight” and “loose” lepton definitions. Each event in the QCD estimate is weighted depending on two variables introduced below. These variables, called  $\epsilon_{\text{eff}}$  and  $\epsilon_{\text{fake}}$ , are estimated in a real lepton-enriched region and a QCD-enriched region.

A QCD-enriched control region is defined orthogonal to the selection region given by the requirements for the analysis (Section 5.3). The control region must be enriched in the QCD multi-jets background, so that the effect of the  $t\bar{t}$  signal and the other backgrounds is reduced. For this analysis, the control region was defined by inverting the missing transverse energy cut and tightening it to  $E_T^{\text{miss}} < 20$  GeV (in both electron and muon channels) and by inverting the transverse mass cut (in the electron channel and in one of the muon channel estimates). An alternate estimate in the muon channel defines the control region by requiring events which have a muon with large impact parameter with respect to the primary vertex. In spite of the difference of control region definitions, both results in the muon channel follow a similar procedure, which is described below.

---

<sup>11</sup>Although it is done by the author in a latter analysis, the  $t\bar{t}$  resonances search at  $\sqrt{s} = 8$  TeV. See Chapter 7.

The main point of the method is to estimate the fraction of events that satisfy the “tight” lepton criteria amongst all events that satisfy the “loose” lepton criteria in the QCD events using the control region, and in true lepton (that is, non-QCD multi-jets) events, using a region enriched in events containing one real lepton. Monte Carlo simulations might be used for the non-QCD region, matching the simulation meta-data to the reconstruction-level leptons to increase the purity. A Tag and Probe method in  $Z \rightarrow \ell\ell$  data events can also define a lepton-enriched region, using the  $Z$  mass window as a selection criteria and using the two leptons to measure the probability of detecting a “tight” lepton as a probe, given that the tag lepton satisfies the “loose” criteria. The latter method is used in this analysis. The two procedures are equivalent, since (see Section 5.6) a Tag And Probe method is used to calculate the lepton selection efficiency in data and use it to correct simulation. In this way, the corrected lepton in simulation has the same selection efficiency as real data.

The fractions determined in the QCD-enriched control region and in the lepton-enriched region can be carried on to the  $t\bar{t}$  selection region and used to define weights for each event, depending on whether it passes the “loose”-only or “tight” criteria.

The number of “loose” events,  $N_{\text{loose}}$ , includes a component,  $N_{\text{loose}}^{\text{fake}}$ , coming from QCD multi-jets events and a component,  $N_{\text{loose}}^{\text{real}}$ , coming from misidentified signal events. And similarly for the number of events satisfying the “tight” criteria, as in Equation 5.4.

$$N_{\text{loose}} = N_{\text{loose}}^{\text{fake}} + N_{\text{loose}}^{\text{real}} \quad (5.3)$$

$$N_{\text{tight}} = N_{\text{tight}}^{\text{fake}} + N_{\text{tight}}^{\text{real}} \quad (5.4)$$

In this method, one wishes to estimate  $N_{\text{tight}}^{\text{fake}}$  in real data, using the selection procedure for the current analysis. We define

$$\epsilon_{\text{eff}} = \frac{N_{\text{tight}}^{\text{real}}}{N_{\text{loose}}^{\text{real}}} \quad (5.5)$$

and

$$\epsilon_{\text{fake}} = \frac{N_{\text{tight}}^{\text{fake}}}{N_{\text{loose}}^{\text{fake}}}, \quad (5.6)$$

which can be used in Equation 5.4 to obtain

$$\begin{pmatrix} N_{\text{loose}} \\ N_{\text{tight}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \epsilon_{\text{fake}} & \epsilon_{\text{eff}} \end{pmatrix} \times \begin{pmatrix} N_{\text{loose}}^{\text{fake}} \\ N_{\text{loose}}^{\text{real}} \end{pmatrix}. \quad (5.7)$$

Note that the number of events that satisfy the tight and loose selections  $N_{\text{loose}}$  and  $N_{\text{tight}}$  in the left-hand side can be found out in data by applying the event selection in data using the “tight”<sup>12</sup> or the slightly-altered “loose” definition for leptons. If the  $\epsilon_{\text{fake}}$  and  $\epsilon_{\text{eff}}$  parameters are known, the unknown values of  $N_{\text{loose}}^{\text{fake}}$  and  $N_{\text{loose}}^{\text{real}}$  could be discovered. Furthermore, these quantities are easily related to the “tight” ones ( $N_{\text{tight}}^{\text{fake}}$  and  $N_{\text{tight}}^{\text{real}}$ ) by Equations 5.5 and 5.6. If we want to calculate  $N_{\text{tight}}^{\text{fake}}$ , which represents the number of QCD events that pass the analysis lepton-definition, in the analysis’ selection region, one can simply calculate the number of events in data that pass the “tight” and “loose” selections and invert the last equation, which results in:

$$\begin{pmatrix} N_{\text{tight}}^{\text{fake}} \\ N_{\text{tight}}^{\text{real}} \end{pmatrix} = \frac{1}{\epsilon_{\text{eff}} - \epsilon_{\text{fake}}} \begin{pmatrix} \epsilon_{\text{fake}}\epsilon_{\text{eff}} & -\epsilon_{\text{fake}} \\ -\epsilon_{\text{fake}}\epsilon_{\text{eff}} & \epsilon_{\text{eff}} \end{pmatrix} \times \begin{pmatrix} N_{\text{loose}} \\ N_{\text{tight}} \end{pmatrix}. \quad (5.8)$$

An interesting element from the last equations is that they are linear on the number of events selected. This implies, that one can divide the data samples in separate sets and  $N_{\text{tight}}^{\text{fake}}$  can be calculated in each one, with its total value for the combined region being the sum of the results in each subset. This feature can be used to apply this equation in an event-by-event basis, creating one separate set for each event. With this setting, each event either passes only the “loose” selection or the event passes both the “loose” and “tight” selections, which would correspond to the  $\begin{pmatrix} N_{\text{loose}} \\ N_{\text{tight}} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  or the  $\begin{pmatrix} N_{\text{loose}} \\ N_{\text{tight}} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  configurations, respectively. The resulting  $N_{\text{tight}}^{\text{fake}}$  value can then be interpreted as a per-event weight to be applied in real data, depending on whether it passes the “tight” selection or only the “loose” selection.

If the data event passes the “tight” and “loose” selections, it is weighted by:

$$w_{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{eff}} - \epsilon_{\text{fake}}} \times (\epsilon_{\text{eff}} - 1), \quad (5.9)$$

---

<sup>12</sup>The one previously mentioned in the event selection for this analysis.

and, events that satisfy only the “loose” criteria, should be weighted by:

$$w_{\text{loose}} = \frac{1}{\epsilon_{\text{eff}} - \epsilon_{\text{fake}}} \times (\epsilon_{\text{fake}} \epsilon_{\text{eff}}). \quad (5.10)$$

The only information necessary for the weights are  $\epsilon_{\text{eff}}$  and  $\epsilon_{\text{fake}}$ , which are estimated in the QCD-enriched and lepton-enriched regions defined previously. The weighting itself is applied on events that pass the selection used in the analysis.

Since the weights are only calculated for each event, the kinematic information of the event can be exploited to parametrise  $\epsilon_{\text{eff}}$  and  $\epsilon_{\text{fake}}$  as a function of a few variables on which they show a large dependency. These variables might vary in different conditions and this parametrisation can improve the shape of the resulting QCD estimate.

These weights are calculated for each event in data, using the standard analysis selection (defined in the Section 5.3). The leptons in the selection are also allowed to pass only a looser criteria (as defined in the current section) if they fail the standard selection with the tighter (default) criteria. Depending on whether the event satisfies the standard selection with the loose or tight lepton criteria, the event is categorised as “loose” or “tight” and the appropriate weight (Equation 5.10 or 5.9, respectively) is associated to it. The histogram for the desired observable (in the current analysis, it would be the jet multiplicity) is filled with the associated event weight and the result is the QCD estimate for such observable for events passing only the tight selection.

Note that in this process, events that are not accepted, if the standard lepton definition is used but are accepted for a looser lepton definition in data, are also included in the QCD estimate, weighted appropriately. According to the equations above, it is expected that the weighted sum of the events satisfying the tight lepton definition and the looser lepton definition is an estimate of the number of QCD multi-jet events that would pass only the tight lepton definition.

## 5.6 Corrections applied in simulation

A set of corrections are applied in simulation to obtain a better agreement between the simulation and data. Some of the corrections are implemented as reweighting procedures, in which the final histograms calculated are weighted based on some characteristics of the event so that some of the selection efficiencies which are different in data and simulation can be taken into account,



or that other differences between data and simulation are reduced, if the simulation setup could not be made exactly as the data configuration. Other correction procedures use statistical pseudo-experiments, which change the kinematics in Monte Carlo.

In some Monte Carlo simulation samples, the events should receive a positive or negative weight according to the MC@NLO procedure described in [55]. Although this involves reweighting, this is not a correction on itself, but a necessary procedure related to how the event generator works. Each sample is also reweighted so that it corresponds to the correct cross section. This is done by normalising a simulation result dividing it by the total number of events generated and multiplying it by the process' production cross section and the data luminosity.

Amongst the reweighting corrections, one of them refers to matching the average number of multiple interactions per bunch crossing  $\langle \mu \rangle$  to data. This is implemented by measuring the average number of interactions per bunch crossing in data,  $\langle \mu \rangle_{\text{data}}$ , and building a normalised histogram with this distribution:  $P_{\mu,\text{data}}$ . A similar distribution is calculated with the average number of simulated interactions for Monte Carlo samples,  $P_{\mu,\text{MC}}$ . With these distributions, a simulated event with an average number of interactions  $\langle \mu \rangle_{\text{MC}}$  receives a weight of  $P_{\mu,\text{data}}(\langle \mu \rangle_{\text{MC}})/P_{\mu,\text{MC}}(\langle \mu \rangle_{\text{MC}})$ , so that the simulated and reweighted  $\langle \mu \rangle$  distributions would match the estimated distribution in data. The  $\langle \mu \rangle$  distribution in data, however, can have different shapes in different running periods and this histogram was calculated separately for different data-taking periods. To account for these different  $\mu$  distributions in data, the Monte Carlo simulation events are separated into subsets with a number of events proportional to the fraction of luminosity in each data period mentioned. Each subset of the Monte Carlo with a fraction of events proportional to a data period range  $X$  is weighted by  $P_{\mu,\text{data period } X}(\langle \mu \rangle_{\text{MC}})/P_{\mu,\text{MC}}(\langle \mu \rangle_{\text{MC}})$ , where  $P_{\mu,\text{data period } X}$  represents the probability distribution of  $\langle \mu \rangle$  in the period range  $X$ . This allows for a reweighting that takes into account features from each data collection period in the  $\langle \mu \rangle$  distribution shape. This procedure is used to get a similar effect from multiple interactions coming from the simulation.

Due to the demand that at least one jet satisfies the  $b$ -tagging criteria, the  $b$ -tagging efficiency and mistag rates could also be a source of discrepancy between data and the simulation, which is why the simulation is reweighted to match the  $b$ -tagging performance in data. The simulation  $b$ -tagging efficiency

and mistag rates are measured for different simulation flavours and different  $\eta$  and  $p_T$  ranges, while in data, different data-driven methods are used to estimate them as a function of  $p_T$  and  $\eta$  (see Section 3.6). The efficiencies are available for each jet and the  $b$ -tagging requirement is applied for any jet in the event (that is, not to a specific jet), so an event-wide weight is built by assuming that there is no correlation on the  $b$ -tagging performance effect between the different jet kinematics. This is implemented by multiplying the scale factor for each jet in a single event weight. The scale factor for each  $b$ -tagged jet is calculated by the ratio of the efficiency of tagging the jet in data and in Monte Carlo simulation. For jets that failed the  $b$ -tagging criteria, the scale factor is calculated as the ratio of the tagging inefficiencies (one minus the efficiency for  $b$ -tagging a jet) in data and Monte Carlo simulation. Details on the scale factors calculated for 2011 ATLAS data are described in [90].

The ratio of efficiencies in data and Monte Carlo is also used to correct for the identification requirements applied in the lepton selection. The document in [80] describes the methods used for muon efficiency estimation in 2010 data, which were used similarly for 2011 data. The electron performance is detailed in [79] for 2010 data as well, but a similar procedure was used for 2011 data. The jet vertex fraction requirement also has a selection efficiency slightly different in simulation and data and the efficiency ratio is used as a scale factor for this correction.

Besides the weighting procedure to correct for efficiency differences in data and simulation, the difference in the resolution and energy scales must also be considered, since the performance in simulation might be slightly different from that in real data collisions. A correction is applied in data for the electron energy scale, using a correction factor measured from data in  $Z$ -boson and  $J/\psi$  decays in electron pairs, and performance studies using  $W$ -boson decays into electrons and neutrinos.

The electron energy resolution is corrected in Monte Carlo simulation by multiplying the electron cluster energy by a pseudo-random sample of a Gaussian distribution with mean one and standard deviation given by  $\sqrt{r_{\text{data}}^2 - r_{\text{MC}}^2}$ , where  $r_{\text{data}}$  is the resolution in data and  $r_{\text{MC}}$  is the resolution in Monte Carlo. The electron performance studies in ATLAS are described in [79] for 2010 data, although the procedure is similar for the 2011 data. The muon transverse momentum is also smeared to take into account the difference in the muon resolution in data and Monte Carlo, using events in real data with two muons in the final state to measure the resolution. Momentum scale correc-

tions are also applied in muons, based on the shift of the  $Z$ -boson peak in measured data. The procedure used to estimate the correction factors is explained in [81] using 2010 data, although the procedure is similar for the 2011 study.

The jet energy scale is also corrected for discrepancies in data and simulation, as described in [87]. The jet energy scale is calculated from energy balance in decays of  $Z$ -boson (which decays into electron or muon pairs) and one extra jet, so that the jet energy calibration can be expressed as a function of the lepton energies. Decays with photons and jets are also used. See Chapter 3 for more details on how these corrections are calculated.

## 5.7 Data to signal and background comparison

After the selection procedure described previously, the number of reconstructed jets can then be analysed, as in Figure 5.1 and compared with data, using the same event selection for all backgrounds. Data-driven techniques are used to estimate the  $W$ +jets background contribution and the QCD jets contribution. Alpgen was used to simulate the  $t\bar{t}$  signal in these plots.

It can be seen that the data and the estimated background agree within systematic uncertainties. This result is going to be used as a first step for the unfolding procedure, detailed in the next sections, in which it is important to subtract the background contribution from data, before moving to the unfolding of the  $t\bar{t}$  signal. The systematic variations on the signal and background were summed in these plots, but for the background subtraction procedure, which is applied in data, only the background systematic uncertainties are taken into account and propagated to the “data - background” estimate, which is detailed in Section 5.10.

More plots showing the kinematic properties of the selected events are available in Appendix A. The numeric yields for each of the signal and backgrounds is given in Table 5.1 for the events that pass the 25 GeV transverse momentum threshold. The systematic uncertainties shown include all variations in the reconstruction process, such as the jet energy scale uncertainty, the  $b$ -tagging mistag rate and efficiency uncertainties, the missing transverse energy uncertainties. More details about the systematic uncertainties will be given in the next section.

As the jet multiplicity shown depends strongly on the transverse momen-

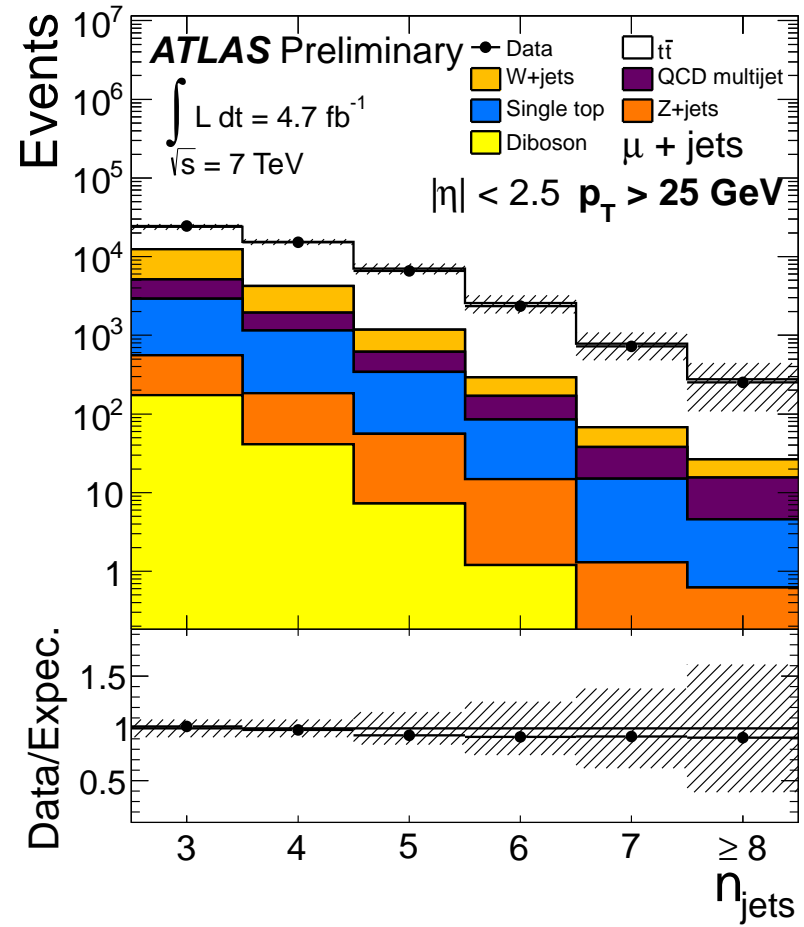
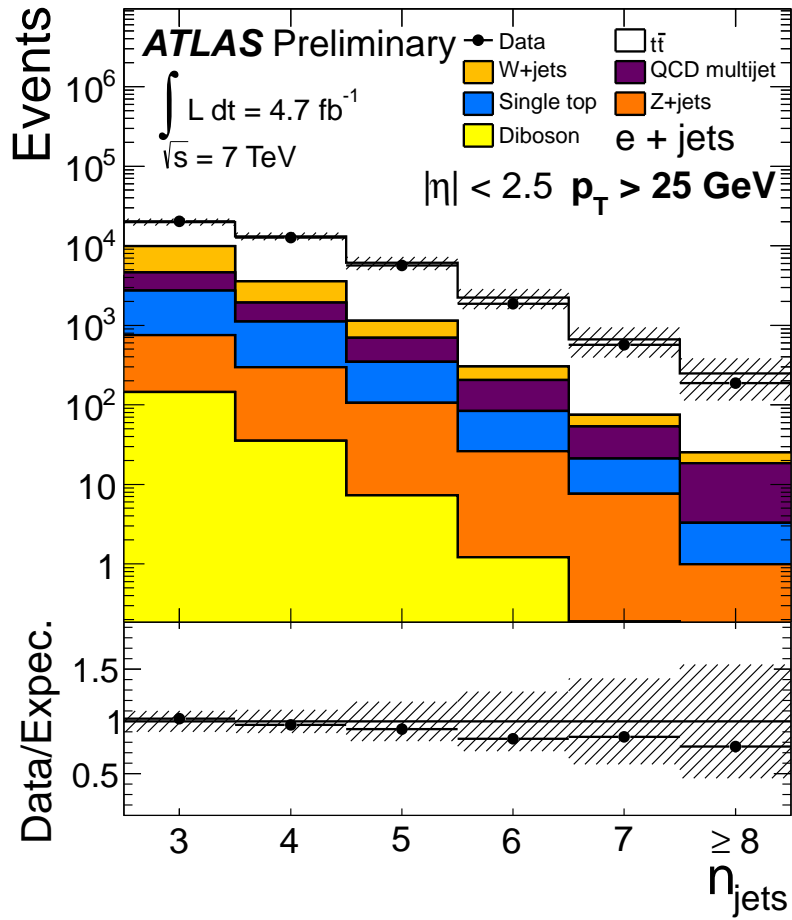


Figure 5.1: Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the  $t\bar{t}$  signal ( $p_T > 25 \text{ GeV}$ ).

Table 5.1: Event yields for data and MC simulation in the electron and muon channels, selected with a 25 GeV jet  $p_T$  threshold. The number of events passing all selection requirements are shown as a function of the reconstructed jet multiplicity ( $n_{\text{jets}}^{\text{reco}}$ ). Alpgen+Herwig is used for the  $t\bar{t}$  simulation and MC expectations are normalised to an integrated luminosity of  $4.7 \text{ fb}^{-1}$ . The uncertainties on the expected values include systematic uncertainties.

Electron channel						
Source	$n_{\text{jets}}^{\text{reco}}=3$	$n_{\text{jets}}^{\text{reco}}=4$	$n_{\text{jets}}^{\text{reco}}=5$	$n_{\text{jets}}^{\text{reco}}=6$	$n_{\text{jets}}^{\text{reco}}=7$	$n_{\text{jets}}^{\text{reco}}\geq 8$
$t\bar{t}$	9897 $\pm$ 1047	9568 $\pm$ 1270	4949 $\pm$ 1102	1922 $\pm$ 619	592 $\pm$ 270	223 $\pm$ 133
W+jets	5333 $\pm$ 1270	1653 $\pm$ 407	441 $\pm$ 137	99 $\pm$ 30	21 $\pm$ 7	7 $\pm$ 2
QCD multijet	1877 $\pm$ 941	818 $\pm$ 410	349 $\pm$ 175	122 $\pm$ 61	32 $\pm$ 16	15 $\pm$ 8
single top	1975 $\pm$ 226	817 $\pm$ 115	242 $\pm$ 48	58 $\pm$ 16	13 $\pm$ 5	2 $\pm$ 2
Z+jets	608 $\pm$ 80	262 $\pm$ 55	99 $\pm$ 23	25 $\pm$ 11	7 $\pm$ 2	1 $\pm$ 2
Diboson	145 $\pm$ 55	35 $\pm$ 14	7 $\pm$ 3	1.21 $\pm$ 0.57	0.19 $\pm$ 0.19	0.11 $\pm$ 0.11
Expectation	19835 $\pm$ 1973	13153 $\pm$ 1467	6088 $\pm$ 1157	2227 $\pm$ 637	666 $\pm$ 274	248 $\pm$ 135
Data	20320	12704	5632	1856	566	188

Muon channel						
Source	$n_{\text{jets}}^{\text{reco}}=3$	$n_{\text{jets}}^{\text{reco}}=4$	$n_{\text{jets}}^{\text{reco}}=5$	$n_{\text{jets}}^{\text{reco}}=6$	$n_{\text{jets}}^{\text{reco}}=7$	$n_{\text{jets}}^{\text{reco}}\geq 8$
$t\bar{t}$	11522 $\pm$ 1191	11156 $\pm$ 1114	5884 $\pm$ 1068	2268 $\pm$ 644	715 $\pm$ 295	250 $\pm$ 168
W+jets	7319 $\pm$ 1473	2296 $\pm$ 477	557 $\pm$ 156	122 $\pm$ 28	29 $\pm$ 7	11 $\pm$ 3
QCD multijet	2201 $\pm$ 451	799 $\pm$ 164	277 $\pm$ 57	85 $\pm$ 17	23 $\pm$ 5	11 $\pm$ 2
single top	2355 $\pm$ 246	965 $\pm$ 129	288 $\pm$ 48	70 $\pm$ 18	14 $\pm$ 5	4 $\pm$ 1
Z+jets	384 $\pm$ 72	142 $\pm$ 30	49 $\pm$ 9	14 $\pm$ 5	1 $\pm$ 3	0.49 $\pm$ 0.69
Diboson	173 $\pm$ 65	41 $\pm$ 16	7 $\pm$ 3	1.20 $\pm$ 0.63	0.03 $\pm$ 0.34	0.13 $\pm$ 0.17
Expectation	23953 $\pm$ 2011	15398 $\pm$ 1295	7061 $\pm$ 1109	2560 $\pm$ 655	782 $\pm$ 300	276 $\pm$ 169
Data	24422	15162	6578	2348	722	252

tum cut applied on the jets before counting them, further comparisons were done with different cuts applied to the jets. Figure 5.2 shows the results using a 40 GeV cut on the jets transverse momentum. Figure 5.3 uses a 60 GeV threshold and Figure 5.4 uses a 80 GeV threshold. A comparison of the Alpgen+Herwig  $t\bar{t}$  simulation with different transverse momentum requirements on the Anti- $k_t$   $R = 0.4$  jets is shown in Figure 5.5.

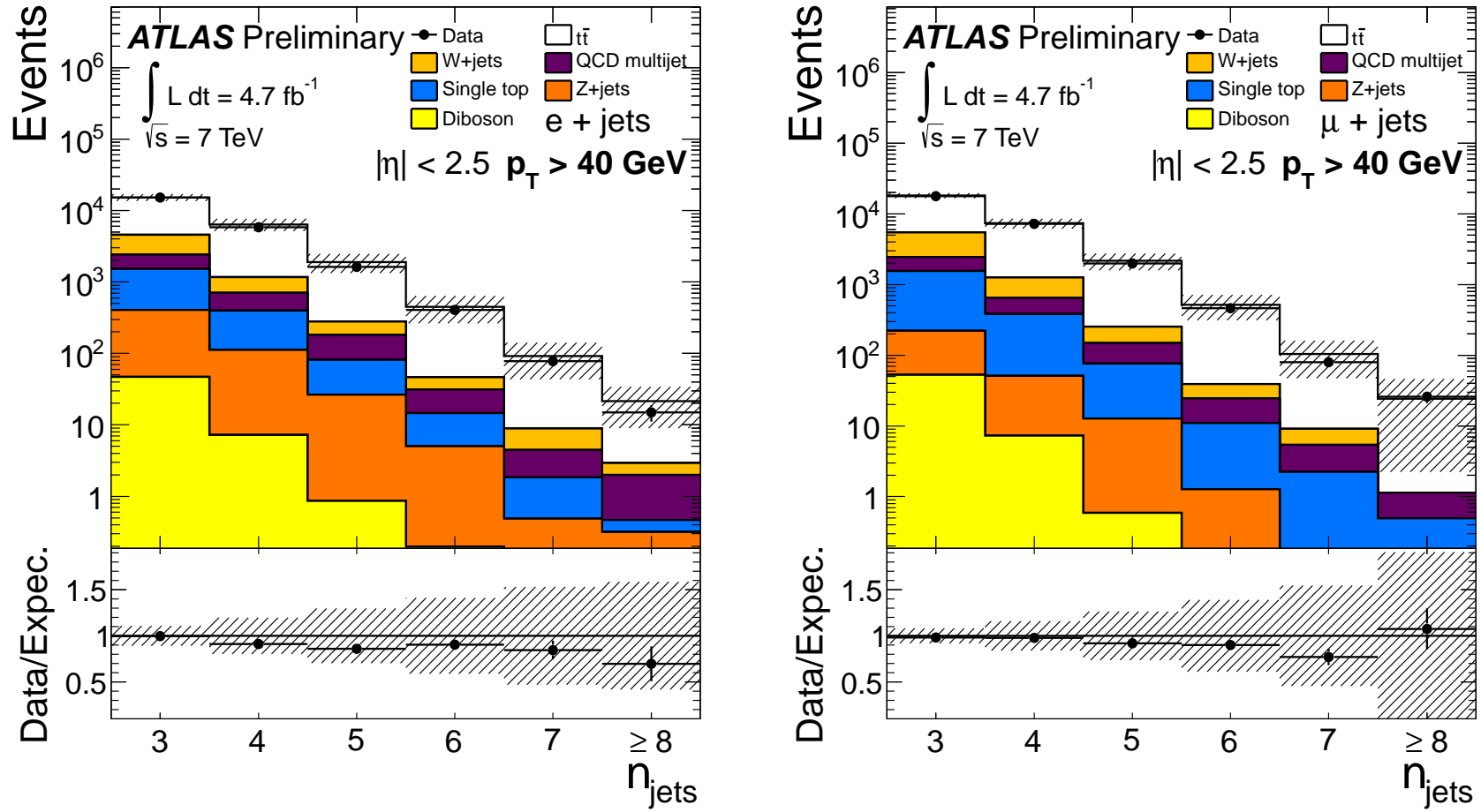
## 5.8 Systematic uncertainties estimate at reconstruction level

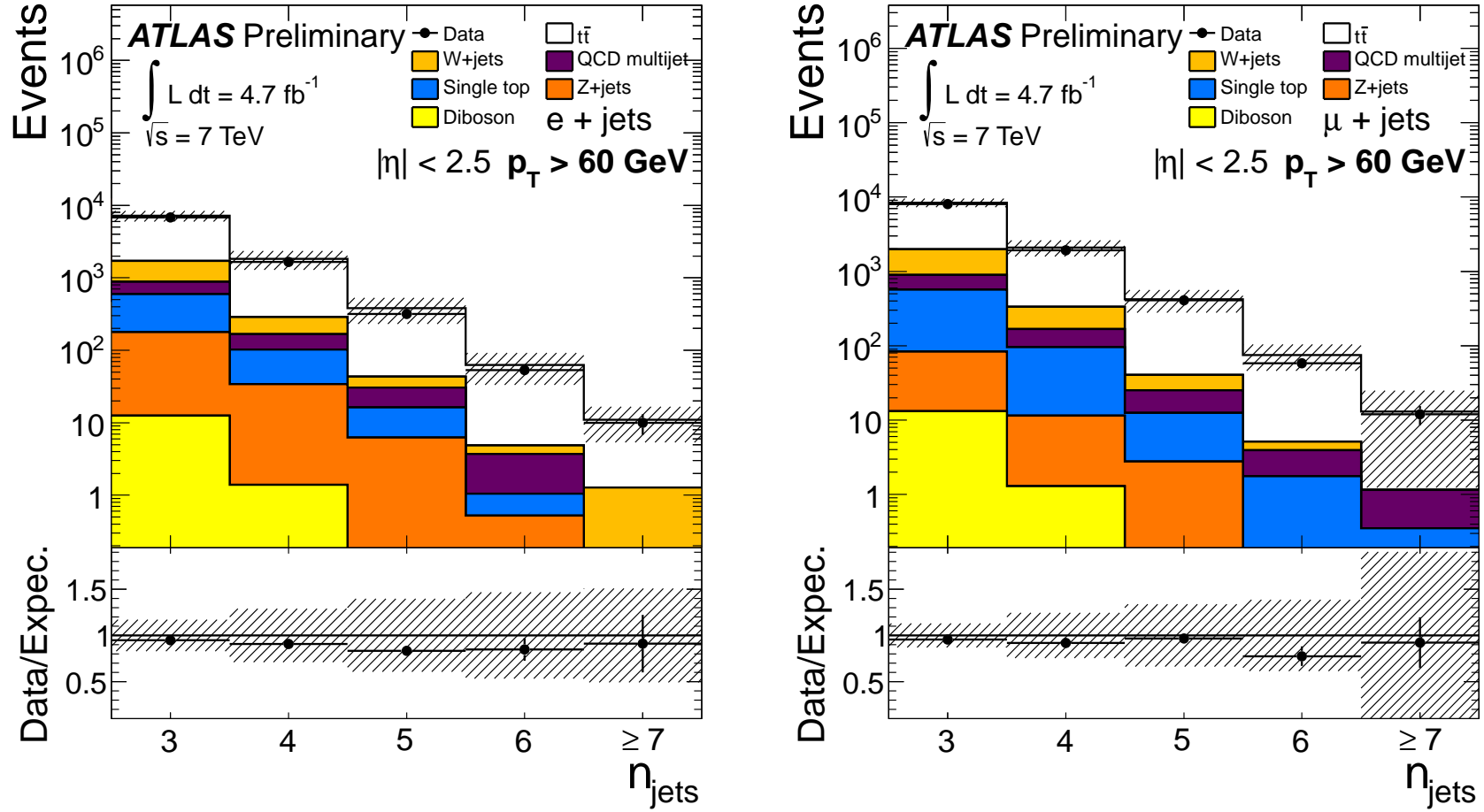
Before correcting for the detector effects, all systematic effects related to the objects reconstruction which affect the selection or the determination of the jet multiplicity are evaluated, by varying each of the parameters used in the calculations. The full list of uncertainties is given in Tables 5.2 and 5.3. A description of the sources of systematic uncertainties is described here.

The jet energy scale is one of the main uncertainties, which has an effect that grows with the jet multiplicity. This can be explained, by noting that, for a higher jet multiplicity, the event has more low transverse momentum jets, which are more sensitive to the minimum transverse momentum requirement. To estimate the uncertainty on the jet energy scale, the jet's four-momentum is varied based on the uncertainties generated by the effect of close-by jets, the effect of multiple proton-proton interactions, and the flavour composition of the jets (light quark versus gluon). For events with more than seven jets, the uncertainty with seven jets was used. An additional  $p_T$  dependent uncertainty was associated to jets that match B-hadrons. The jet reconstruction efficiency was measured as the fraction of jets reconstructed from tracks that match a calorimeter jet and the difference observed was taken as a jet reconstruction efficiency uncertainty [85], which was applied in this measurement by randomly removing a fraction of the jets in the simulation events accordingly.

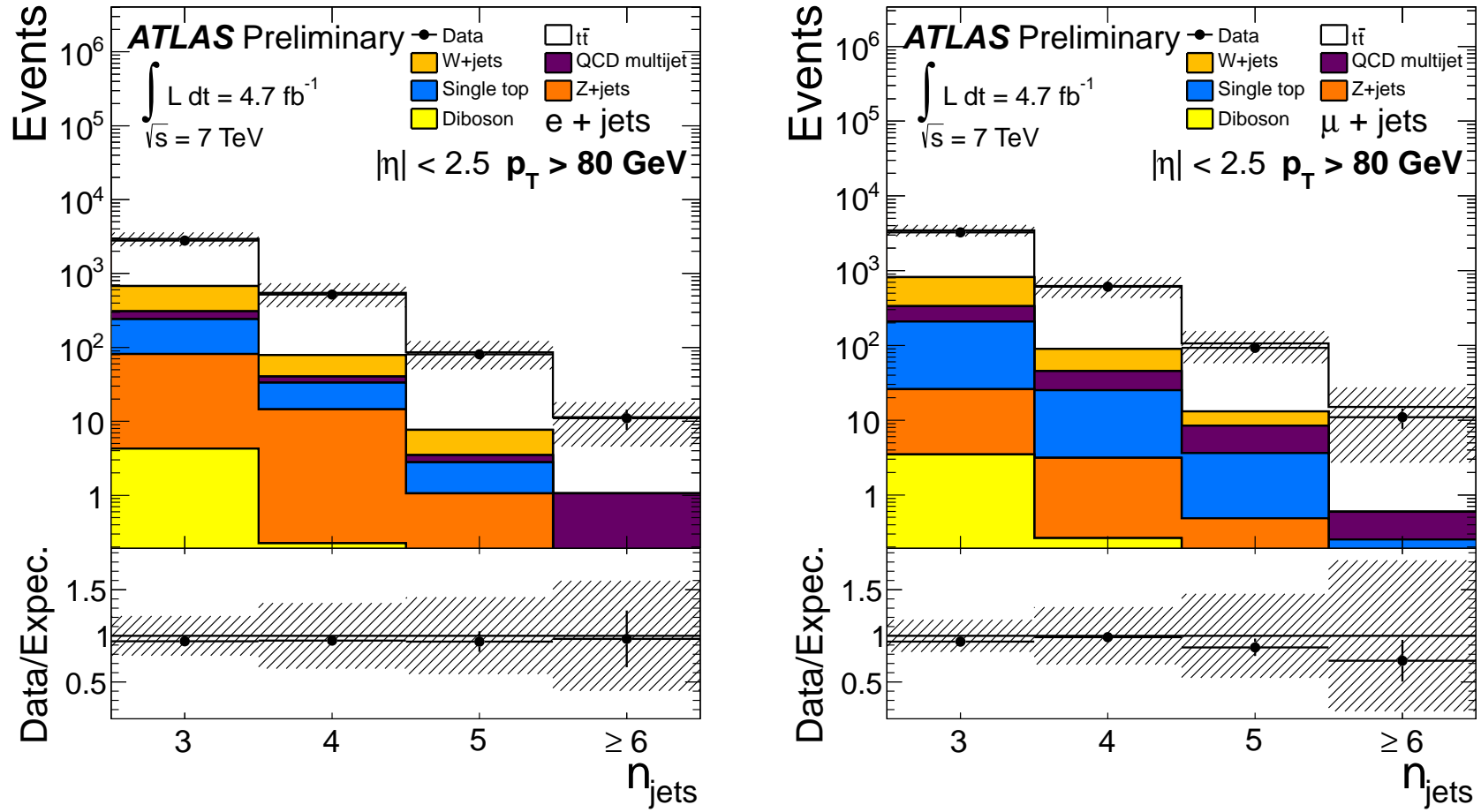
Jets were also smeared according to the jet energy resolution uncertainty, after checking that there is an agreement in this quantity between data and simulation [85, 89]. No nominal correction was applied for the jet energy resolution, since studies [89] show good agreement between data and simulation.

Another one of the main uncertainties, is the  $b$ -tagging performance, which was measured in data and simulation. The differences in the data and simulation  $b$ -tagging efficiencies and mistag rates were corrected in the simulation by applying a scale factor to all events [92]. The uncertainty in the measure-

Figure 5.2: Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the  $t\bar{t}$  signal ( $p_T > 40 \text{ GeV}$ ).

Figure 5.3: Jet multiplicity in the electron (left) and muon (right) channels using Alpgen simulation for the  $t\bar{t}$  signal ( $p_T > 60 \text{ GeV}$ ).



Figure 5.4: Jet multiplicity in the electron (left) and muon (right) channel using Alpgen simulation for the  $t\bar{t}$  signal ( $p_T > 80 \text{ GeV}$ ).

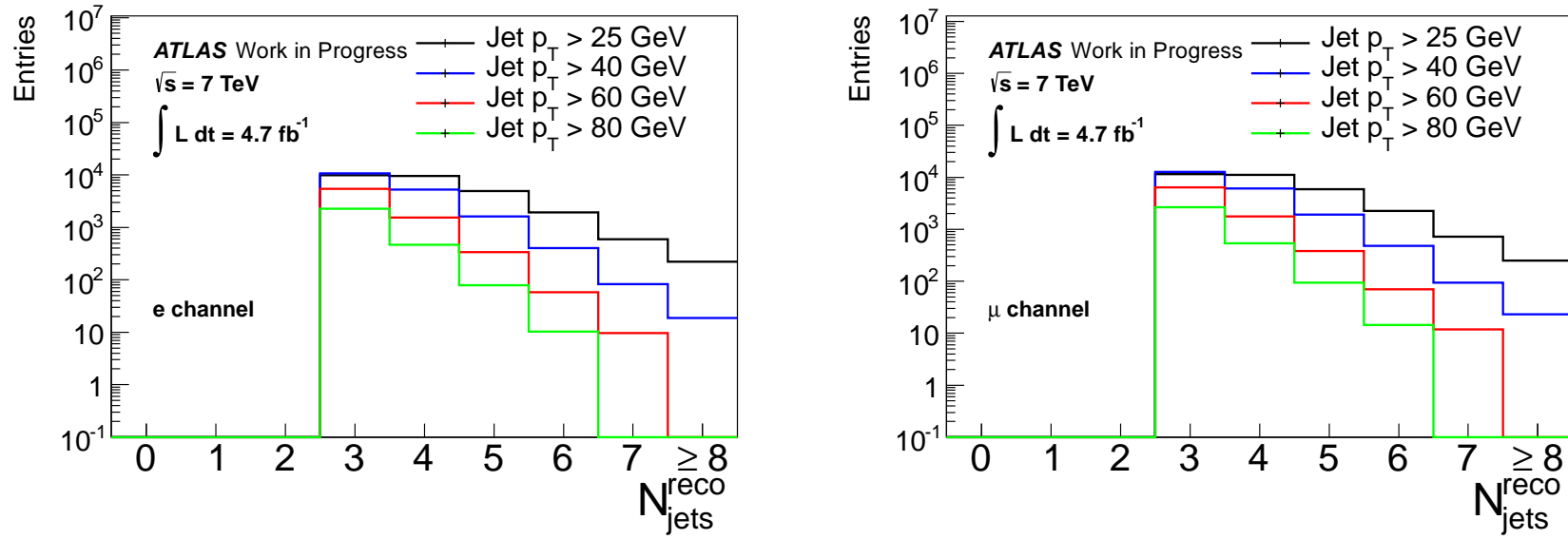


Figure 5.5: Jet multiplicity in the electron (left) and muon (right) channels using Alpgen+Herwig simulation for the  $t\bar{t}$  signal with different Anti- $k_t$  jet transverse momentum cuts applied. In this figure, for the 60 GeV plot, the 7 jet bin represents events with  $\geq 7$  jets, and in the 80 GeV plot, the 6 jet bin represents events with  $\geq 6$  jets.

ment of the  $b$ -tagging performance was propagated to the scale factors and its effect in the final observable was estimated by varying the scale factors used in simulation accordingly. The effect of this uncertainty is significant in the low jet multiplicity bins, but it does not grow as much as the jet energy scale uncertainty with the number of jets in the event. The scale factors used and their uncertainties can be seen in Section 3.6, in Figure 3.14.

To mitigate the effect of multiple proton-proton interactions, the jets under consideration are required to have a jet vertex fraction greater than 0.75 in absolute value [95]. A scale factor was applied based on the efficiency ratio in data and simulation events and the uncertainty in the efficiency measurement was propagated to the scale factor, as mentioned in Section 3.5. The leptons' trigger, reconstruction and identification efficiencies were also measured in simulation and in data and scale factors were also derived with the appropriate uncertainties (see Section 3.3 and Section 3.4). The efficiencies were measured in data through  $Z$  and  $W$  boson decays.

The missing transverse energy is measured by summing all corrected lepton and jet energies and demanding that the transverse energy is conserved. Calorimeter cells not associated to reconstructed objects with  $p_T > 20$  GeV, have their energies added in the missing transverse energy “CellOut” component. “Soft” jets, that is, cells from jets with  $p_T > 7$  GeV and  $p_T < 20$  GeV, and the “CellOut” components are varied by 6.6% to estimate the effect of the multiple proton-proton interactions in the selection. This number was calculated by studying the dependency of the missing transverse energy on the multiple particle interactions [95]. Consult Section 3.7 for more information.

The Alpgen  $t\bar{t}$  prediction has an uncertainty from the choice of the CTEQ6L1 Parton Distribution Function (PDF) [38], which was evaluated by using the MSTW PDF set at leading-order with 68% Confidence Level [41, 102] to reweight the  $t\bar{t}$  sample. The systematic uncertainty related to the PDF choice was calculated by including the difference in the nominal value caused by the choice of the PDF, as well as adding in quadrature the difference between the nominal value and the results when using all eigenvector sets from the MSTW PDF [102]. The uncertainty due to the parton shower modelling was also estimated by comparing the results obtained with Alpgen+Herwig and Alpgen+Pythia  $t\bar{t}$  simulations. The ISR/FSR variations between the Alpgen+Pythia central yields and the Alpgen+Pythia yields with  $\alpha_S$  increased and decreased were also used to estimate the extra radiation impact in the results. The Powheg+Pythia and the Alpgen+Pythia  $t\bar{t}$  samples were com-

pared to include the systematic uncertainty related to the difference between fixed order matrix element calculations and associated matching schemes.

The W+jets charge asymmetry measurement also has an uncertainty associated to it, from statistical uncertainties on the data-driven measurement and uncertainties from the simulation-dependent part of the method, including lepton and jets reconstruction, charge mis-identification, Monte Carlo generators, backgrounds and Parton Distribution Function uncertainties. The heavy flavour fraction estimate includes, besides simulation uncertainties, a 25% uncertainty when extrapolating the results from the 2-jet bin to higher jet multiplicity. The W+jets Monte Carlo simulation also includes an uncertainty from the choice of the renormalisation and factorisation scales (estimated by varying the `iqopt3` parameter<sup>13</sup> in Alpgen) and from generator cuts (estimated by varying the `ptjmin` parameter<sup>14</sup> in Alpgen).

The integrated luminosity in data was measured using van der Meer [103] scans and it was used to normalise many simulation samples accordingly. Its uncertainty was found to be 3.9%<sup>15</sup> [103]. The single top production cross section uncertainties were taken to be 4% in the  $t$ -channel, 4% in the  $s$ -channel and 8% in the  $Wt$ -channel from approximate NNLO calculations. The diboson production cross section uncertainty was taken to be 5%. For Z+jets, 4% added in quadrature with 24% per jet was taken for the theoretical cross section uncertainty.

The QCD multi-jet background uncertainty can be estimated in the muon channel from the shape difference between the two methods used for the data-driven estimate. In the electron channel, the missing transverse energy selection requirement was varied between 15 GeV and 25 GeV for the control region. The normalisation uncertainty was taken to be 50% in the electron channel and 20% in the muon channel as a result of comparing the Matrix Method estimates with other methods.

---

<sup>13</sup>This changes the method in which the scale is defined in Alpgen, by multiplying the standard scale by  $\sqrt{m_W^2 + p_{T,W}^2}$ . See [44] for more information.

<sup>14</sup>This changes the minimum transverse momentum cut for light jets in Alpgen. See [44] for more information.

<sup>15</sup>An uncertainty of 3.9% was used instead of 3.7%, as in the reference, due to the higher uncertainties in the second half of the 2011 data taking.

## 5.9 Unfolding the effect of the detector

When reconstructing the jet, lepton and missing transverse energy quantities, the detector changes the physical observables in many ways. An unfolding procedure is necessary to measure the actual cross section as a function of the jet multiplicity and correct the effect of the detector. The reference used as the particle level result is obtained after applying only the transverse momentum, isolation and  $\eta$  requirements on the Monte Carlo simulation metadata. The Alpgen+Herwig sample, used in the previous data to simulation comparison plots is used as a reference sample for the unfolding procedure. The systematic uncertainty associated with using a particular reference sample is estimated. The propagation of uncertainties through the unfolding method is described in Section 5.10.

The particle-level selection demands one electron or muon and no other lepton with the same pseudo-rapidity and transverse momentum requirements as the reconstruction-level selection. The electron's four-momentum is summed with photons around it in a  $\Delta R < 0.1$  region, to simulate the effect of radiation emitted when interacting with the detector. The sum of final state particles' transverse momentum around the lepton (excluding its own momentum and disregarding neutrinos) with  $p_T > 500$  MeV and  $\Delta R < 0.3$  is required to be smaller than 2 GeV, to simulate the acceptance effect of the isolation cuts applied to the leptons in the reconstruction-level selection<sup>16</sup>. The same overlap removal criteria are also demanded for the leptons, as in the reconstruction-level selection, to simulate their acceptance. The jets are built at the particle-level selection, using the Anti- $k_t$  algorithm with the same  $R = 0.4$  configuration applied to the simulation meta-data and the same transverse momentum and pseudorapidity ranges are demanded. The missing transverse energy at particle-level is calculated summing the simulation meta-data for particles that interact with the detector and taking their negative transverse momentum. The same missing transverse energy and transverse mass requirements are applied in the simulation. At least one of the particle-level jets is required to satisfy  $\Delta R < 0.3$  between the jet axis and a  $B$ -hadron, to simulate the  $b$ -tagging criteria demanded in the reconstruction-level selection.

The corrections in this method are expressed as:

---

<sup>16</sup>For electrons, photons in the  $\Delta R < 0.1$  region are excluded from the particle-level isolation calculation, since they were used to build the electron's four-momentum, simulating its interaction with the detector.

$$\begin{aligned}
N_{\text{jets}}^{\text{part}}(i) &= \mathcal{U}_{j \rightarrow i} [N_{\text{jets}}^{\text{reco}}(j) - N_{\text{jets}}^{\text{bkg}}(j)] \\
&= \frac{1}{f_{\text{reco}}(i)} \sum_{j=3}^8 f(i, j) (1 - f_{\text{np3}}(j)) (1 - f'_{\text{fakes}}(j)) \times \\
&\quad [N_{\text{jets}}^{\text{reco}}(j) - N_{\text{jets}}^{\text{bkg}}(j)]
\end{aligned} \tag{5.11}$$

where  $N_{\text{jets}}^{\text{reco}}(j)$  represents the number of entries at reconstruction-level jet multiplicity  $j$  and  $N_{\text{jets}}^{\text{part}}(i)$  represents the number of entries at particle-level jet multiplicity  $i$ . Whenever the index is not explicitly mentioned and the lower case variables  $n_{\text{jets}}^{\text{reco}}$  and  $n_{\text{jets}}^{\text{part}}$  are used, they will be taken to mean the jet multiplicity values for a single event. The operator  $\mathcal{U}_{j \rightarrow i}$  represents the unfolding process to be applied in background subtracted input and it is defined by the equation above.

Starting with the reconstructed jet multiplicity spectrum  $N_{\text{jets}}^{\text{reco}}(j)$ , the background as estimated in Section 5.2 is subtracted through the  $N_{\text{jets}}^{\text{bkg}}(j)$  term.

The next steps are represented by the unfolding operator  $\mathcal{U}_{j \rightarrow i}$ . In the steps that follow describing the unfolding operator, the jet multiplicity requirement at the reconstruction-level selection or the particle-level selection are not taken into account unless explicitly mentioned. The reason for this is that this selection requirement is analysed independently. That means that, for what follows, “reconstructed events” will be used to refer to events that pass the reconstruction-level selection, with no requirement on  $n_{\text{jets}}^{\text{reco}}$ , and “particle-level events” refers to events that pass the particle-level selection, with no requirement on  $n_{\text{jets}}^{\text{part}}$ . The abbreviation “R” is used for events that pass the reconstruction-level selection, regardless of the  $n_{\text{jets}}^{\text{reco}}$  requirement and “P” for events that pass the particle-level selection, regardless of the  $n_{\text{jets}}^{\text{part}}$  requirement.

The unfolding steps in the unfolding operator start (from right to left in Equation 5.11) with the removal of events that were reconstructed but that fail the particle level cuts except the jet multiplicity requirement. The fraction of events reconstructed with  $n_{\text{jets}}^{\text{reco}} \geq 3$  that failed the particle level cuts is defined as  $f'_{\text{fakes}}$ :

$$f'_{\text{fakes}} = \frac{\text{Number of events in “R” with } n_{\text{jets}}^{\text{reco}} \geq 3, \text{ but not in “P”}}{\text{Number of events in “R” with } n_{\text{jets}}^{\text{reco}} \geq 3}. \tag{5.12}$$

This factor estimates the fraction of fake  $t\bar{t}$  events reconstructed in the signal sample used. The value calculated for  $1 - f'_{\text{fakes}}$  is shown in Figure 5.6. It is important to mention that the jet multiplicity is a consequence of the jet requirements applied, that is, the acceptance cuts. This correction factor includes acceptance effects in the reconstruction procedure related to the jets, but not the particle-level effects. Note, as well, that Figure 3.14 shows the  $b$ -tagging selection efficiency, which is no more than 20%, which has a major impact in the acceptance. The  $f'_{\text{fakes}}$  factor also contains a contribution from electron and muon misidentification and the jet vertex fraction selection requirement.

Since the particle-level jet multiplicity cut was disregarded in the previous step, it must be taken into account separately. The  $f_{\text{np3}}$  factor is the fraction of events that pass the reconstruction-level and particle-level selections, but failed the requirement  $n_{\text{jets}}^{\text{part}} \geq 3$ :

$$f_{\text{np3}} = \frac{\text{Number of events in "R" (with } n_{\text{jets}}^{\text{reco}} \geq 3) \text{ and in "P", but fail } n_{\text{jets}}^{\text{part}} \geq 3}{\text{Number of events in "R" (with } n_{\text{jets}}^{\text{reco}} \geq 3) \text{ and in "P"}}. \quad (5.13)$$

The multiplication by  $1 - f_{\text{np3}}$  removes events that migrated from particle-level bins 0, 1 and 2 to reconstruction-level bins  $\geq 3$ . The value calculated for  $1 - f_{\text{np3}}$  is shown in Figure 5.7.

A migration correction from reconstruction-level to particle-level is applied as a matrix multiplication. Each element in the migration matrix,  $f$ , is the conditional probability that an event was at particle-level bin  $j$ , given that it was reconstructed at bin  $i$ , that is:

$$f(i, j) = \frac{\text{Number of events in "R" and "P" with } n_{\text{jets}}^{\text{part}} = j \text{ and } n_{\text{jets}}^{\text{reco}} = i}{\text{Number of events in "R" with } n_{\text{jets}}^{\text{reco}} = i}. \quad (5.14)$$

The factor  $f$  is already an *unsmearing* factor which can be directly multiplied in Equation 5.11. The factor  $f$  is calculated by counting the number of events in the  $n_{\text{jets}}^{\text{reco}}$  and  $n_{\text{jets}}^{\text{part}}$  bins and normalising the matrix by the reconstruction-level bins, so that  $\sum_{j=3}^8 f(i, j) = 1 \forall i \in [3, 8]$ . The migration factor  $f$  for the selection with a jet  $p_T$  cut at 25 GeV is shown in Figure 5.8.

Finally, a correction is applied to include events that exist at particle level, but that were lost during the reconstruction procedure. The fraction of events that exist at reconstruction level, given that they can be found at particle

level, is given by  $f_{\text{reco}}$ :

$$f_{\text{reco}} = \frac{\text{Number of events in "R" and in "P"}}{\text{Number of events in "P"}}. \quad (5.15)$$

The value calculated for  $f_{\text{reco}}$  is shown in Figure 5.9. Note that this effect is corrected after the unsmearing performed by the  $f$  matrix and after the out-of-acceptance correction between reconstruction and particle-level performed by the  $f'_{\text{fakes}}$  and  $f_{\text{np3}}$ . As a consequence, this final step only corrects for the acceptance difference for the events accepted by the particle-level selection.

The knowledge of  $n_{\text{jets}}^{\text{reco}}$  for an event before the event selection is not trivial, since an overlap removal is done between jets and electrons, which removes one or zero jets, changing the value of  $n_{\text{jets}}^{\text{reco}}$  by one or zero. However, it is not obvious which electron is to be selected (if any) and  $n_{\text{jets}}^{\text{reco}}$  cannot be pre-calculated exactly. This method uses the number of reconstructed jets,  $n_{\text{jets}}^{\text{reco}}$ , only at the stages in which the reconstruction-level selection was fulfilled, which avoids this difficulty.

The effect of each of the corrections is shown in Figure 5.10 for the Alpgen top-antitop sample, assuming perfect background subtraction. In this case, the unfolded result matches perfectly the particle-level result as expected, because the same sample was used to derive the correction factors and to apply them. The figure shows that the method is able to recover the original particle-level result from a reconstruction-level measurement.

The equivalent plots for the selection jet  $p_T$  cut at 40, 60 and 80 GeV are available in Section 5.12.

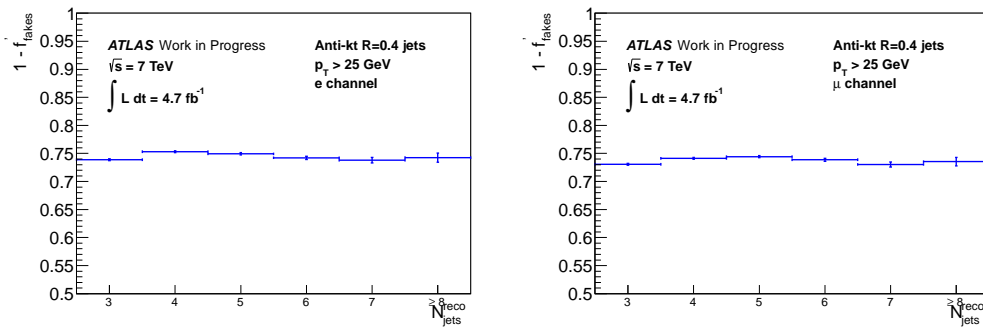


Figure 5.6: The  $1 - f'_{\text{fakes}}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown.



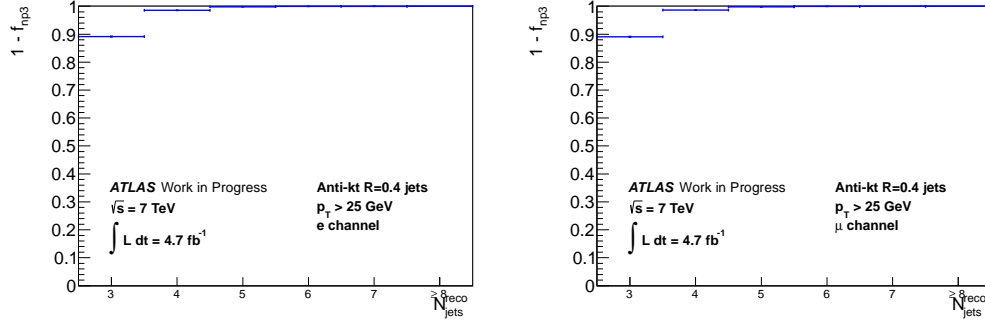


Figure 5.7: The  $1 - f_{np3}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown.

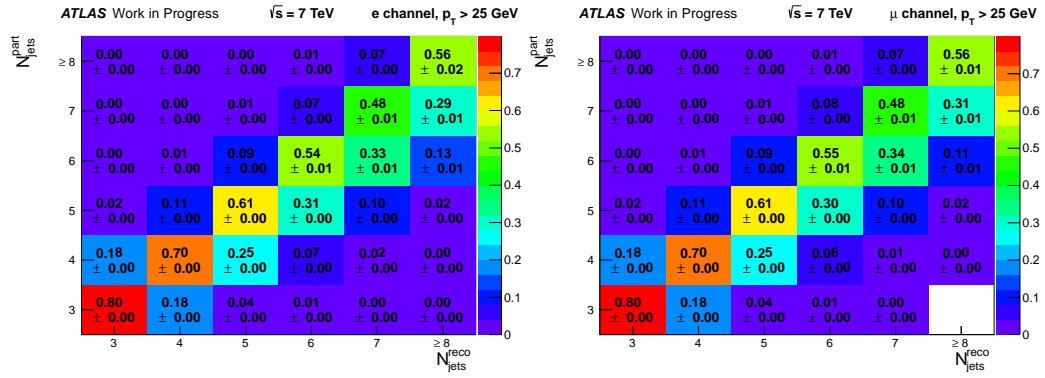


Figure 5.8: The migration matrix using the Alpgen  $t\bar{t}$  signal sample with a selection using a jet  $p_T$  cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown.

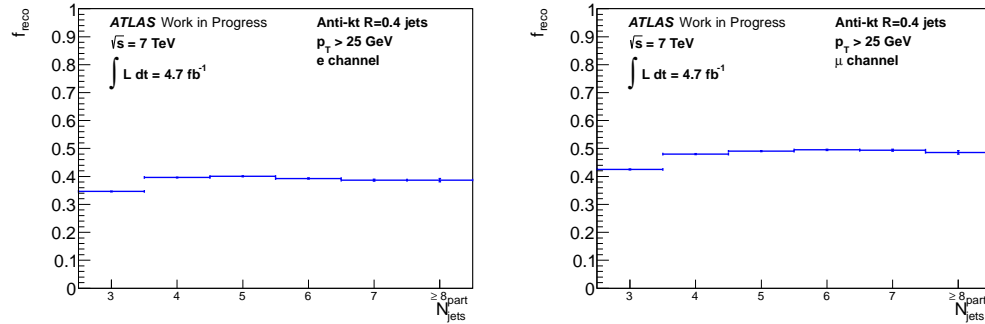


Figure 5.9: The  $f_{reco}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 25 GeV. The results for the electron (left) and muon (right) channels are shown.

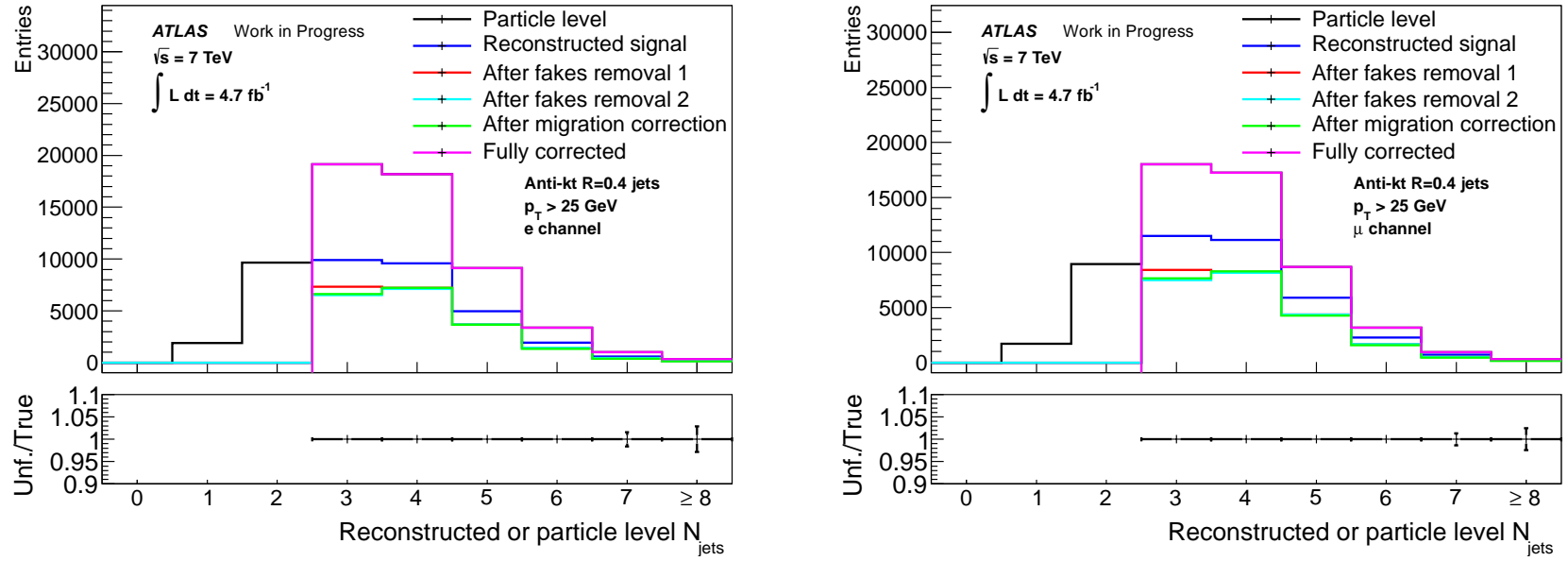


Figure 5.10: The closure test using the Alpgen  $t\bar{t}$  signal sample for input and corrections with a jet  $p_T$  cut at 25 GeV for the selection. The results for the electron (left) and muon (right) channels are shown.

## 5.10 Propagation of systematic uncertainties through the unfolding procedure

There are different sets of uncertainties that should be propagated through the method described previously. The background is subtracted from data using a bin-by-bin subtraction of the data histogram and the reconstruction systematic variations of the backgrounds are propagated to the background-subtracted data sample<sup>17</sup>. This results in a set of systematic uncertainties which need to be propagated through the  $\mathcal{U}_{j \rightarrow i}$  operator described previously.

The systematic uncertainty in the reconstruction of the  $t\bar{t}$  signal sample used to estimate the unfolding correction factors should be considered separately, since the  $n_{\text{jets}}^{\text{reco}}$  variable calculated in this sample is not actually used in the unfolding procedure except indirectly through the usage of the correction factors. It is proposed to calculate the fraction of the uncertainty in each  $N_{\text{jets}}^{\text{reco}}$  bin for the  $t\bar{t}$  sample used in deriving the correction factors and apply this fraction as a variation to the background-subtracted data sample. Note as well that this systematic variation is fully anti-correlated with the background reconstruction systematic variation in the background-subtracted data, since a positive variation in the background sample causes a negative variation in the background-subtracted sample. Therefore, the background-subtracted data is multiplied by the positive fractional variation of the simulation  $t\bar{t}$  sample to get a negative systematic variation to be added in the equivalent background systematic variation.

With the previous set of systematic uncertainties in the real data sample, a set of variations should be found in  $I(j) = N_{\text{jets}}^{\text{reco}}(j) - N_{\text{jets}}^{\text{bkg}}(j)$  from real data, which incorporate the systematic effect of the reconstruction in the signal and background estimates. It is not assumed that the propagation of uncertainties through the unfolding procedure  $N_{\text{jets}}^{\text{part}}(i) = \mathcal{U}_{j \rightarrow i}[I(j)]$  leads to an unfolded uncertainty in  $N_{\text{jets}}^{\text{part}}$  with the same distribution as the uncertainty in  $I$ . The method implemented considers each source of systematic uncertainty  $s$ , at each bin  $j$  in the background-subtracted data sample  $I$  separately.  $\delta(s, j)$  is used to refer to the absolute systematic variation caused by  $s$ , in bin  $j$ , so that the estimated sample with this variation is  $I_s(s, j) = I(j) + \delta(s, j)$  for each bin  $j$ . The systematic variation  $\delta(s, j)$  represents a Gaussian standard deviation due to the systematic uncertainty source  $s$  and it can be calculated from the nominal

---

<sup>17</sup>A positive systematic in the background would become a negative systematic variation in the new sample, but the absolute value of the variations would not change.

background-subtracted data,  $I(j)$ , and the corresponding reconstruction-level systematic variation,  $I_s(s, j)$ , which is calculated for each uncertainty source, as described in Section 5.8. Pseudo-experiments are performed to establish the effect of the unfolding procedure  $\mathcal{U}_{j \rightarrow i}$  on the source  $s$ .

For each source  $s$ , a number  $\mathcal{N}$  of pseudo-random samples of a Gaussian with mean zero and standard deviation one are taken and they are referred to as  $\alpha(s, m)$ , for integers  $m \in [1, \mathcal{N}]$ . A pseudo-random systematic variation is defined as:

$$I_{ps}(s, m, j) \triangleq I(j) + \alpha(s, m)\delta(s, j), \quad (5.16)$$

which has mean  $I(j)$  for each bin  $j$  and standard deviation  $\delta(s, j)$ , as desired. Each  $I_{ps}(s, m, j)$  is unfolded into  $\mathcal{U}_{j \rightarrow i}[I_{ps}(s, m, j)]$ , which has a mean given by:

$$\xi(s, i) \triangleq \frac{1}{\mathcal{N}} \sum_{m=1}^{\mathcal{N}} \mathcal{U}_{j \rightarrow i}[I_{ps}(s, m, j)]. \quad (5.17)$$

The measure  $\gamma(s, i)$  of the systematic effect of the source  $s$  after the unfolding procedure is defined as the sample variance of  $\mathcal{U}_{j \rightarrow i}[I_{ps}(s, m, j)]$ :

$$\gamma(s, i) \triangleq \frac{1}{\mathcal{N} - 1} \sqrt{\sum_{m=1}^{\mathcal{N}} (\mathcal{U}_{j \rightarrow i}[I_{ps}(s, m, j)] - \xi(s, i))^2}. \quad (5.18)$$

This procedure gives us a  $\gamma(s, i)$  for each bin  $i$  and each systematic variation  $s$ , which is used as an estimate of this systematic variation after the unfolding procedure. In this analysis, the number of pseudo-experiments was taken to be  $\mathcal{N} = 1000$  due to computational limitations.

The procedure above is also implemented using the  $t\bar{t}$  simulation as an input. The  $t\bar{t}$  simulation reconstruction-level histogram is represented as  $I_{t\bar{t}}(j)$ , similarly to the  $I(j)$  histogram for background-subtracted data. All reconstruction-level systematics in the  $t\bar{t}$  simulation can be propagated into an unfolded systematic variation in the same way as it was described previously, resulting in a measure of the systematic effect  $s$  given by  $\gamma_{t\bar{t}}(s, i)$  for the bin  $i$  of the unfolded  $t\bar{t}$   $N_{\text{jets}}^{\text{part}, t\bar{t}}(i) = \mathcal{U}_{j \rightarrow i}[I_{t\bar{t}}(j)]$ . The total effect of the reconstruction-level systematic variation  $s$  in the background-subtracted and unfolded data is given by:

$$\gamma_{\text{reco}}(s, i) \triangleq \gamma(s, i) - N_{\text{jets}}^{\text{part}}(i) \cdot \frac{\gamma_{t\bar{t}}(s, i)}{N_{\text{jets}}^{\text{part}, t\bar{t}}(i)}, \quad (5.19)$$

while the nominal background-subtracted data is given by  $N_{\text{jets}}^{\text{part}}(i) = \mathcal{U}_{j \rightarrow i}[I(j)]$ . Note that the subtraction is used, since the background and  $t\bar{t}$  systematic uncertainties are fully anti-correlated.

Another set of systematic uncertainties on the  $t\bar{t}$  modelling had a different treatment. These uncertainties include the Parton Distribution Functions (PDFs), the initial state radiation and final state radiation (ISR/FSR), the parton shower modelling, the Monte Carlo generator systematics and the unfolding factors' statistical uncertainty<sup>18</sup>. The signal distribution at reconstruction-level  $I_{\text{model}}(c, j)$  for each different configuration  $c$  and bin  $j$  is unfolded as  $\mathcal{U}_{j \rightarrow i}[I_{\text{model}}(c, j)]$  and the systematic effect  $\gamma_{\text{model}}(c, j)$  is defined as:

$$\gamma_{\text{model}}(c, i) \triangleq |\mathcal{U}_{j \rightarrow i}[I_{\text{model}}(c, j)] - I_{\text{particle}}(c, i)|, \quad (5.20)$$

where  $I_{\text{particle}}(c, i)$  is the particle-level jet multiplicity value for configuration  $c$ , in bin  $i$ . The total systematic uncertainty is calculated as:

$$\gamma_{\text{total}}(i) \triangleq \sqrt{\sum_s \gamma_{\text{reco}}^2(s, i) + \sum_c \gamma_{\text{model}}^2(c, i)}, \quad (5.21)$$

where it is implicit that for systematic uncertainties that contain asymmetric variations, the maximum (in absolute value) variation is used and symmetrised and does not enter the sum.

The systematic uncertainties on the unfolded distributions are shown in Tables 5.4 and 5.5. The values shown are percentages of the unfolded data.

## 5.11 Results at particle level and discussion

The unfolded jet multiplicity distributions are shown in Figures 5.11, 5.12, 5.13 and 5.14. These plots only show the final number of entries after the correction implemented. As a final step, the corrected number of entries was divided by the integrated luminosity to estimate the fiducial cross section in Figures 5.15, 5.16, 5.17 and 5.18. The black line shows the unfolded data and the shaded band indicates the propagated systematics. The green line shows the Alpgen  $t\bar{t}$  signal for comparison. It can be seen that the unfolded data is

---

<sup>18</sup>The samples used for each configuration are described in Section 5.2. The statistical uncertainty in the unfolding factors was taken into consideration, by applying the procedure described in this paragraph to a statistically independent sample to the  $t\bar{t}$  dataset used to derive the unfolding correction factors, but which was generated in the same way. The MC@NLO sample was used to derive the Monte Carlo generator systematics.

compatible with the Alpgen  $t\bar{t}$  signal. The plots in a logarithm scale for the Y axis are also shown in Figures 5.19, 5.20, 5.21 and 5.22 for the corrected number of events. The final fiducial cross section is shown in Figures 5.23, 5.24, 5.25 and 5.26.

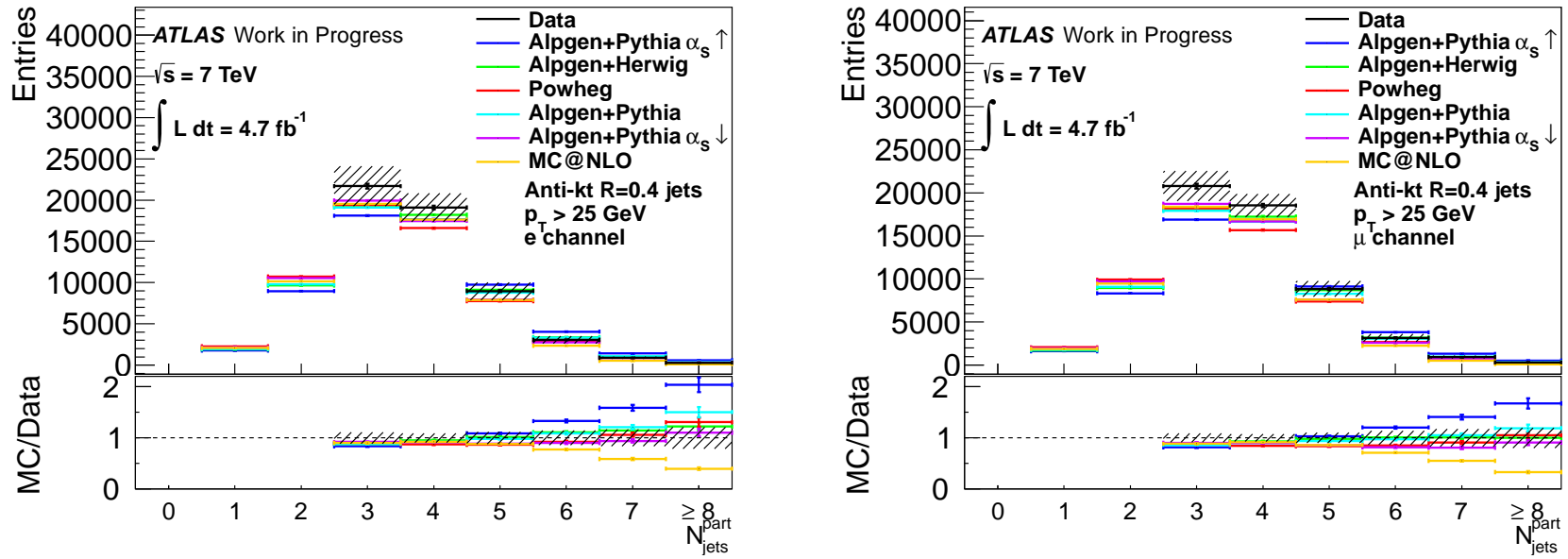


Figure 5.11: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 25 GeV.

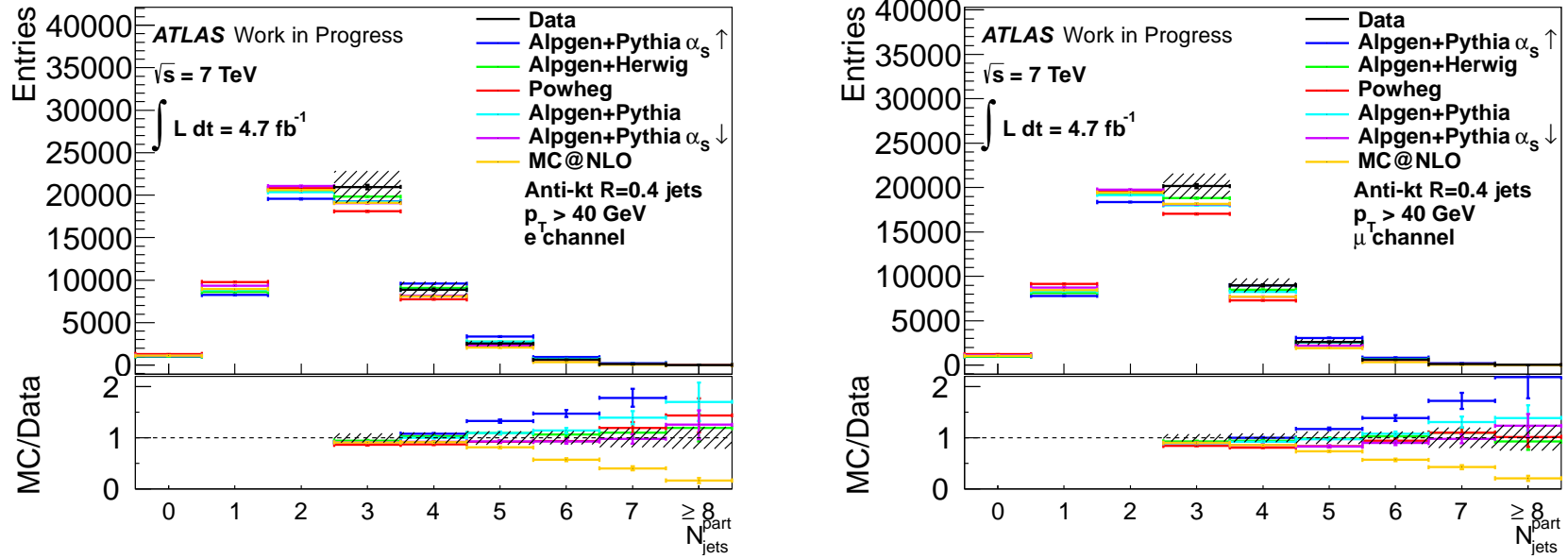


Figure 5.12: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 40 GeV.



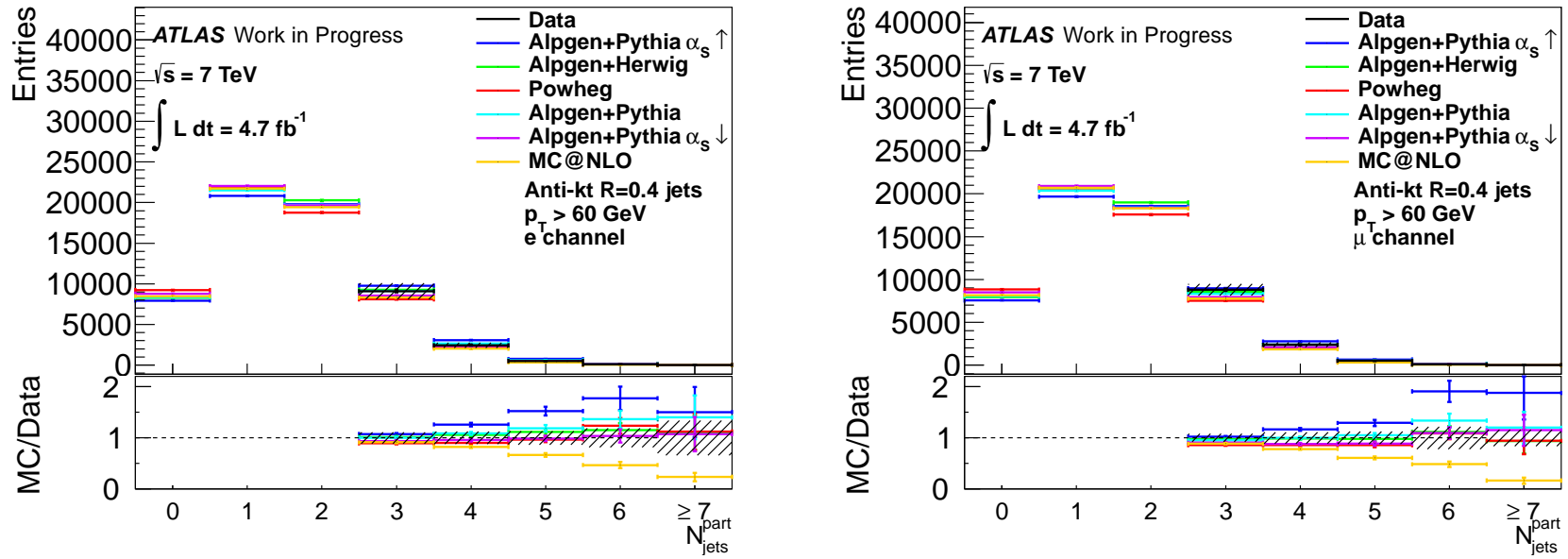


Figure 5.13: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 60 GeV.

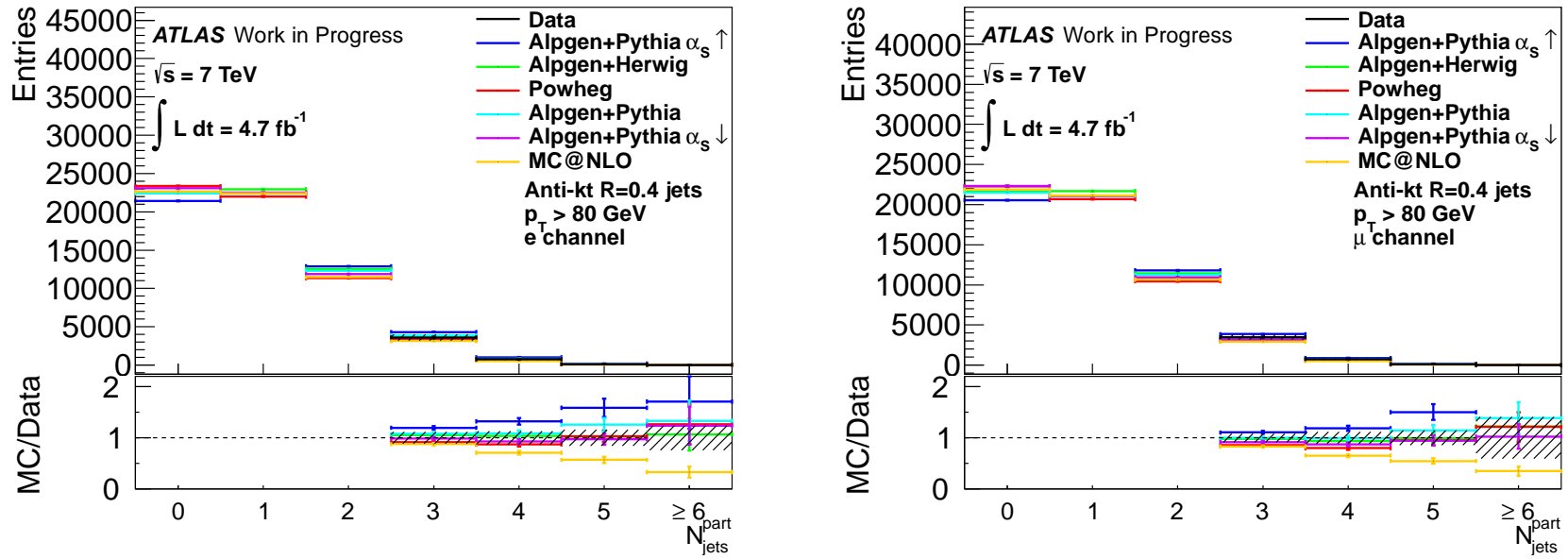


Figure 5.14: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 80 GeV.

Table 5.2: Uncertainties on event yields at reconstruction level in the electron channel, selected with a 25 GeV jet  $p_T$  threshold. Alpgen is used for the  $t\bar{t}$  simulation. The uncertainties are shown as a percentage of the expected  $t\bar{t}$  signal.

Syst.	Integral	$N_{\text{jets}}^{\text{reco}} = 3$	$N_{\text{jets}}^{\text{reco}} = 4$	$N_{\text{jets}}^{\text{reco}} = 5$	$N_{\text{jets}}^{\text{reco}} = 6$	$N_{\text{jets}}^{\text{reco}} = 7$	$N_{\text{jets}}^{\text{reco}} \geq 8$
b-tagging eff.	3.46	3.08	3.69	3.96	3.96	4.12	3.38
b-tagging c tag rate	0.42	0.44	0.41	0.42	0.43	0.34	0.50
mistag rate	0.41	0.40	0.40	0.41	0.51	0.47	0.47
Cell Out	0.05	0.03	0.06	0.04	0.01	0.24	0.65
Electron energy scale	0.02	0.02	0.02	0.02	0.06	0.05	0.09
Electron energy res.	0.33	0.30	0.34	0.40	0.35	0.28	0.31
Jet efficiency	0.06	0.06	0.00	0.14	0.16	0.16	0.31
Jet energy res.	0.52	0.04	0.15	1.52	2.42	5.19	4.58
Jet energy scale	3.95	1.27	3.15	7.82	14.57	20.51	26.22
JVF	1.06	0.75	1.14	1.46	1.79	2.12	2.42
Lepton Identification	1.64	1.40	1.79	1.92	1.99	2.03	2.01
Lepton Reconstruction	0.65	0.55	0.71	0.77	0.80	0.81	0.80
Muon mom. res. (MS)	0.01	0.01	0.01	0.00	0.00	0.00	0.00
Muon mom. res. (ID)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon mom. scale	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Pile up	0.04	0.03	0.05	0.03	0.05	0.15	0.49
Luminosity	3.20	2.85	3.41	3.62	3.73	3.78	3.79
W+jets shape	0.27	0.24	0.05	0.97	0.00	0.00	0.00
W+jets norm.	1.92	2.66	1.38	1.31	0.80	0.58	0.49
W+jets bb+25%	1.99	2.90	1.51	0.81	0.60	0.48	0.31
W+jets bb+cc vs. c	0.79	1.04	0.68	0.40	0.46	0.45	0.26
W+jets c+25%	1.29	2.11	0.75	0.37	0.17	0.10	0.10
W+jets bb+cc+c vs. light	2.97	4.46	2.09	1.21	0.77	0.57	0.40
ISR/FSR	4.29	3.10	0.43	7.60	17.72	28.54	37.29
MC generator	6.09	0.92	7.86	13.74	15.32	18.20	22.38
Parton shower	1.64	1.35	1.50	1.52	2.02	6.79	17.65
DD QCD norm.	3.80	4.73	3.11	2.86	2.73	2.42	3.06
DD QCD eff. shape	0.34	0.35	0.34	0.32	0.30	0.28	0.36
DD QCD fake shape	3.42	4.25	2.73	2.50	2.50	4.22	2.86
Diboson norm.	0.15	0.25	0.09	0.04	0.02	0.01	0.02
Single top norm.	0.57	0.77	0.48	0.31	0.20	0.16	0.07
PDF syst.	0.21	0.23	0.14	0.12	0.30	0.76	1.93
All syst.	13.92	10.82	11.49	19.17	28.72	39.75	61.88
Stat. MC	0.33	0.45	0.44	0.57	0.84	1.46	2.37
Stat. Data	0.48	0.72	0.86	1.23	1.93	3.57	5.53

Table 5.3: Uncertainties on event yields at reconstruction level in the muon channel, selected with a 25 GeV jet  $p_T$  threshold. Alpgen is used for the  $t\bar{t}$  simulation. The uncertainties are shown as a percentage of the expected  $t\bar{t}$  signal.

Syst.	Integral	$N_{\text{jets}}^{\text{reco}} = 3$	$N_{\text{jets}}^{\text{reco}} = 4$	$N_{\text{jets}}^{\text{reco}} = 5$	$N_{\text{jets}}^{\text{reco}} = 6$	$N_{\text{jets}}^{\text{reco}} = 7$	$N_{\text{jets}}^{\text{reco}} \geq 8$
b-tagging eff.	3.25	2.72	3.60	3.94	3.96	4.22	3.65
b-tagging c tag rate	0.38	0.36	0.39	0.38	0.45	0.41	0.58
mistag rate	0.35	0.35	0.31	0.38	0.44	0.44	0.35
Cell Out	0.13	0.10	0.18	0.11	0.11	0.12	0.09
Electron energy scale	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Electron energy res.	0.01	0.01	0.02	0.00	0.01	0.01	0.00
Jet efficiency	0.03	0.00	0.03	0.11	0.05	0.14	0.28
Jet energy res.	0.51	0.31	0.28	0.48	1.88	5.03	6.71
Jet energy scale	3.55	1.47	2.34	6.99	13.00	20.09	29.17
JVF	1.05	0.73	1.13	1.48	1.82	2.14	2.52
Lepton Identification	0.54	0.45	0.59	0.65	0.68	0.69	0.68
Lepton Reconstruction	0.24	0.20	0.27	0.29	0.30	0.31	0.30
Muon mom. res. (MS)	0.02	0.02	0.01	0.02	0.00	0.04	0.05
Muon mom. res. (ID)	0.01	0.00	0.01	0.01	0.02	0.02	0.02
Muon mom. scale	0.18	0.15	0.20	0.23	0.18	0.21	0.18
Pile up	0.08	0.06	0.11	0.04	0.08	0.08	0.13
Luminosity	3.09	2.71	3.32	3.59	3.71	3.75	3.74
W+jets shape	0.27	0.23	0.02	1.07	0.00	0.00	0.00
W+jets norm.	1.82	2.54	1.27	1.18	0.71	0.56	0.60
W+jets bb+25%	1.87	2.59	1.55	0.85	0.47	0.45	0.47
W+jets bb+cc vs. c	0.55	0.71	0.51	0.26	0.19	0.32	0.38
W+jets c+25%	1.92	3.03	1.21	0.57	0.31	0.15	0.12
W+jets bb+cc+c vs. light	2.65	3.87	1.97	1.09	0.58	0.43	0.41
ISR/FSR	4.48	3.04	0.45	8.52	19.34	28.60	43.87
MC generator	4.09	1.07	5.13	9.21	8.54	11.74	13.36
Parton shower	1.10	0.41	1.31	1.70	0.24	6.59	26.21
DD QCD norm.	1.36	1.84	1.04	0.78	0.66	0.59	0.79
DD QCD shape estimate	1.09	1.32	1.00	0.75	0.65	0.60	0.75
Diboson norm.	0.15	0.25	0.09	0.04	0.02	0.00	0.02
Single top norm.	0.57	0.76	0.48	0.31	0.21	0.13	0.11
PDF syst.	0.21	0.23	0.14	0.12	0.30	0.83	2.10
All syst.	11.07	8.40	8.41	15.70	25.60	38.31	61.05
Stat. MC	0.25	0.40	0.38	0.47	0.67	1.14	2.02
Stat. Data	0.44	0.65	0.80	1.15	1.89	3.44	5.74

Table 5.4: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The  $p_T$  cut on the jets is 25 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} = 7$	$N_{\text{jets}}^{\text{part}} \geq 8$
b-tagging eff.	2.44	2.36	2.46	2.54	2.53	2.53	2.37
b-tagging c tag rate	0.10	0.04	0.11	0.17	0.21	0.22	0.26
mistag rate	0.08	0.04	0.08	0.11	0.16	0.22	0.25
Cell Out	0.03	0.03	0.04	0.03	0.01	0.09	0.14
Electron energy scale	0.01	0.00	0.02	0.01	0.02	0.01	0.04
Electron energy res.	0.19	0.15	0.21	0.22	0.23	0.21	0.19
Jet efficiency	0.05	0.04	0.01	0.12	0.17	0.19	0.28
Jet energy res.	0.47	0.06	0.40	0.61	2.04	4.26	3.61
Jet energy scale	2.11	0.76	1.16	4.13	7.75	11.47	14.63
JVF	0.78	0.65	0.77	0.94	1.12	1.31	1.52
Lepton Identification	1.17	1.09	1.19	1.25	1.28	1.30	1.32
Lepton Reconstruction	0.44	0.41	0.45	0.48	0.49	0.50	0.50
Muon mom. res. (MS)	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Muon mom. res. (ID)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon mom. scale	0.01	0.01	0.00	0.01	0.00	0.00	0.00
Pile up	0.04	0.02	0.06	0.03	0.03	0.07	0.15
Luminosity	2.12	1.98	2.17	2.27	2.33	2.36	2.40
W+jets shape	0.40	0.38	0.27	0.81	0.34	0.09	0.02
W+jets norm.	1.69	2.43	1.38	1.03	0.75	0.54	0.45
W+jets bb+25%	1.79	2.70	1.48	0.82	0.55	0.41	0.31
W+jets bb+cc vs. c	0.71	0.99	0.62	0.40	0.37	0.34	0.27
W+jets c+25%	1.13	1.86	0.84	0.40	0.21	0.11	0.09
W+jets bb+cc+c vs. light	2.62	4.05	2.11	1.18	0.76	0.53	0.40
DD QCD norm.	3.63	4.67	3.14	2.62	2.66	2.47	2.90
DD QCD eff. shape	0.29	0.34	0.28	0.25	0.24	0.23	0.22
DD QCD fake shape	3.22	4.16	2.76	2.31	2.39	2.29	2.87
Unfolding syst. (other MC gen.)	2.96	5.25	1.22	1.12	1.89	4.35	10.63
Unfolding syst. (same sample)	1.11	1.13	1.04	1.23	1.06	0.71	2.68
ISR/FSR	3.44	0.69	3.84	7.18	8.32	6.56	6.33
Parton shower	2.38	0.01	3.48	5.10	4.27	2.59	3.50
PDF syst.	3.06	2.88	3.02	3.30	3.59	3.84	3.76
All syst.	5.87	11.09	8.97	11.80	14.11	16.18	21.14
Stat.	0.71	1.36	1.08	1.32	1.95	3.35	6.44

Table 5.5: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The  $p_T$  cut on the jets is 25 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} = 7$	$N_{\text{jets}}^{\text{part}} \geq 8$
b-tagging eff.	2.74	2.74	2.75	2.72	2.69	2.71	2.57
b-tagging c tag rate	0.16	0.10	0.17	0.21	0.25	0.27	0.33
mistag rate	0.09	0.04	0.10	0.14	0.20	0.24	0.22
Cell Out	0.09	0.08	0.11	0.08	0.07	0.04	0.02
Electron energy scale	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Electron energy res.	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jet efficiency	0.03	0.00	0.01	0.08	0.08	0.15	0.25
Jet energy res.	0.62	0.59	0.27	0.44	1.70	4.08	6.44
Jet energy scale	2.07	1.09	0.84	3.66	7.20	11.23	16.39
JVF	0.78	0.65	0.76	0.93	1.10	1.29	1.49
Lepton Identification	0.38	0.35	0.39	0.41	0.42	0.43	0.43
Lepton Reconstruction	0.18	0.17	0.18	0.19	0.19	0.20	0.20
Muon mom. res. (MS)	0.01	0.01	0.00	0.01	0.00	0.01	0.01
Muon mom. res. (ID)	0.00	0.00	0.00	0.00	0.01	0.00	0.01
Muon mom. scale	0.23	0.23	0.24	0.25	0.21	0.22	0.13
Pile up	0.05	0.03	0.08	0.05	0.07	0.03	0.07
Luminosity	2.06	1.91	2.10	2.21	2.27	2.31	2.32
W+jets shape	0.38	0.34	0.24	0.79	0.32	0.08	0.01
W+jets norm.	1.62	2.43	1.31	0.88	0.61	0.45	0.41
W+jets bb+25%	1.67	2.50	1.44	0.80	0.46	0.35	0.32
W+jets bb+cc vs. c	0.48	0.69	0.44	0.25	0.17	0.20	0.23
W+jets c+25%	1.61	2.66	1.24	0.58	0.29	0.15	0.10
W+jets bb+cc+c vs. light	2.34	3.66	1.94	1.03	0.57	0.37	0.30
DD QCD norm.	1.23	1.80	1.03	0.68	0.53	0.46	0.52
DD QCD shape	0.96	1.29	0.87	0.62	0.49	0.44	0.48
Unfolding syst. (other MC gen.)	2.66	2.31	1.79	3.85	5.63	6.17	2.63
Unfolding syst. (same sample)	1.68	1.55	1.71	1.80	1.74	2.10	2.72
ISR/FSR	2.97	1.28	2.63	5.69	7.03	7.55	4.77
Parton shower	1.79	0.76	2.10	2.86	3.41	2.90	2.02
PDF syst.	3.23	2.96	3.24	3.47	3.75	4.36	4.73
All syst.	4.60	8.38	7.26	10.10	13.42	16.90	19.73
Stat.	0.64	1.24	0.98	1.16	1.67	2.80	5.34

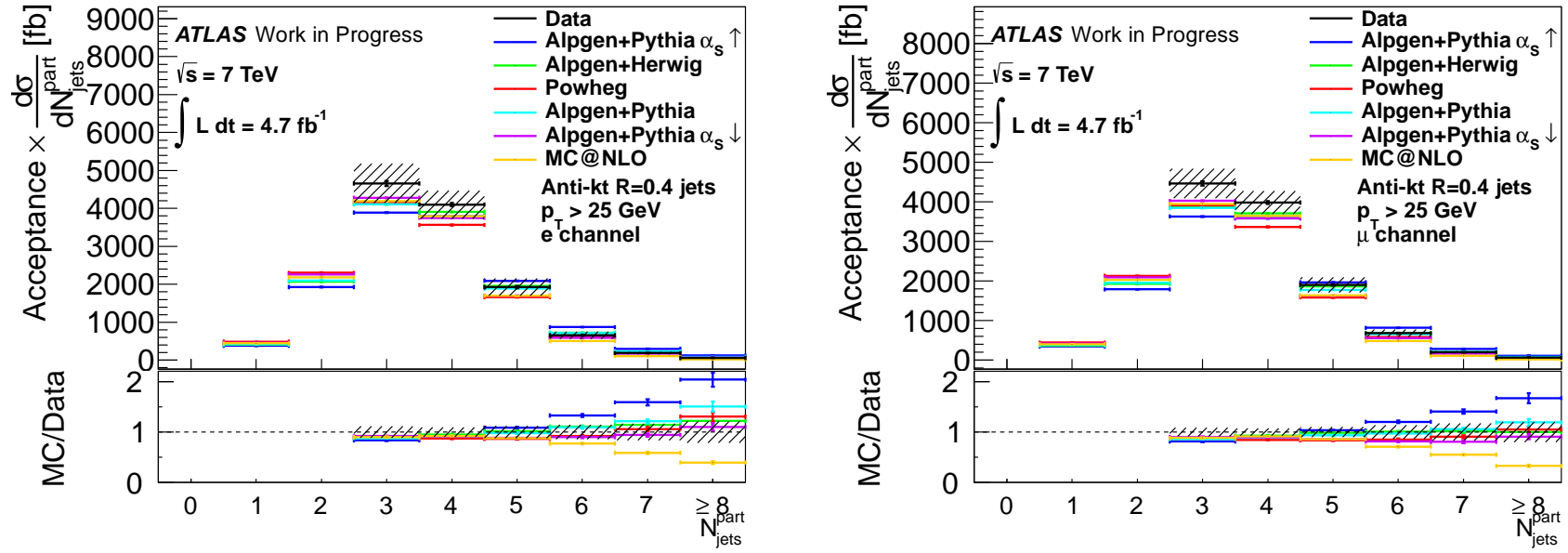


Figure 5.15: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 25 GeV.

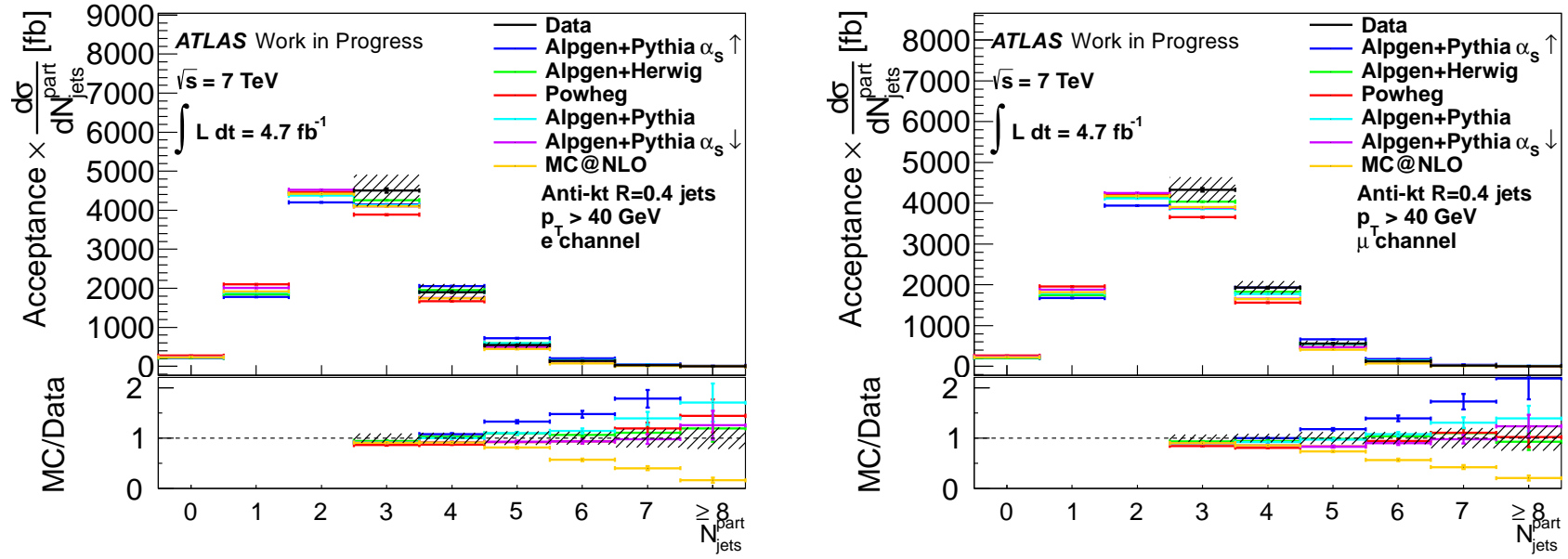


Figure 5.16: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 40 GeV.



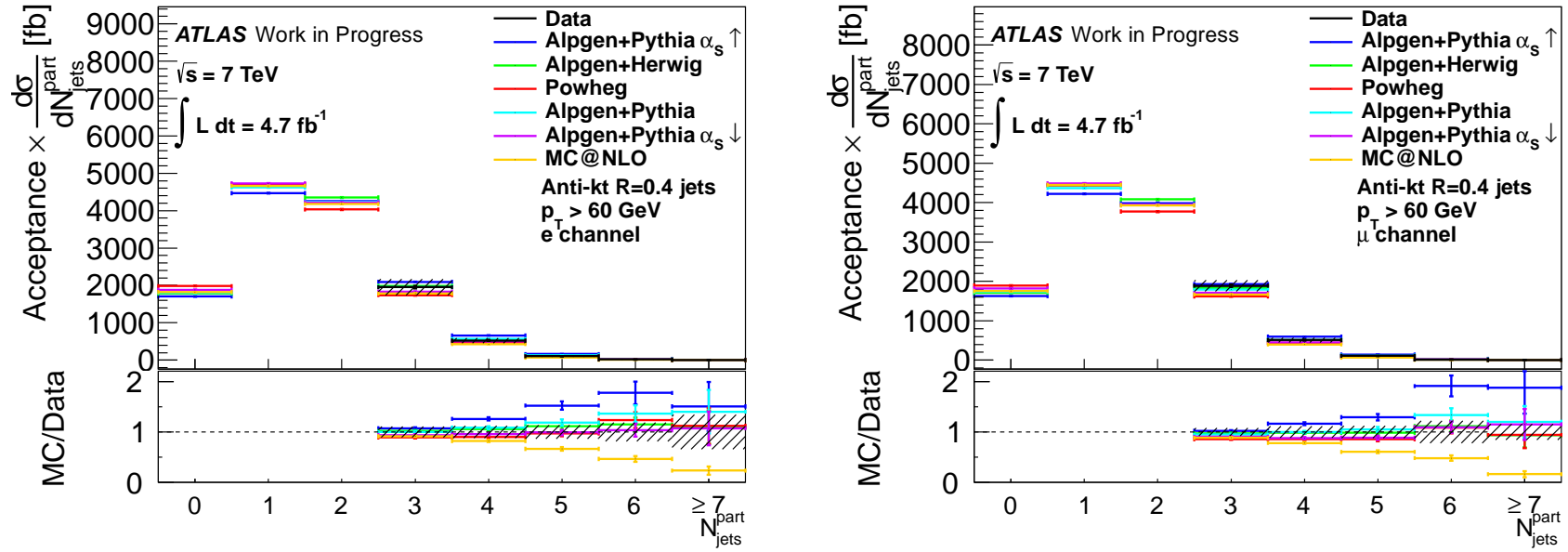


Figure 5.17: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 60 GeV.

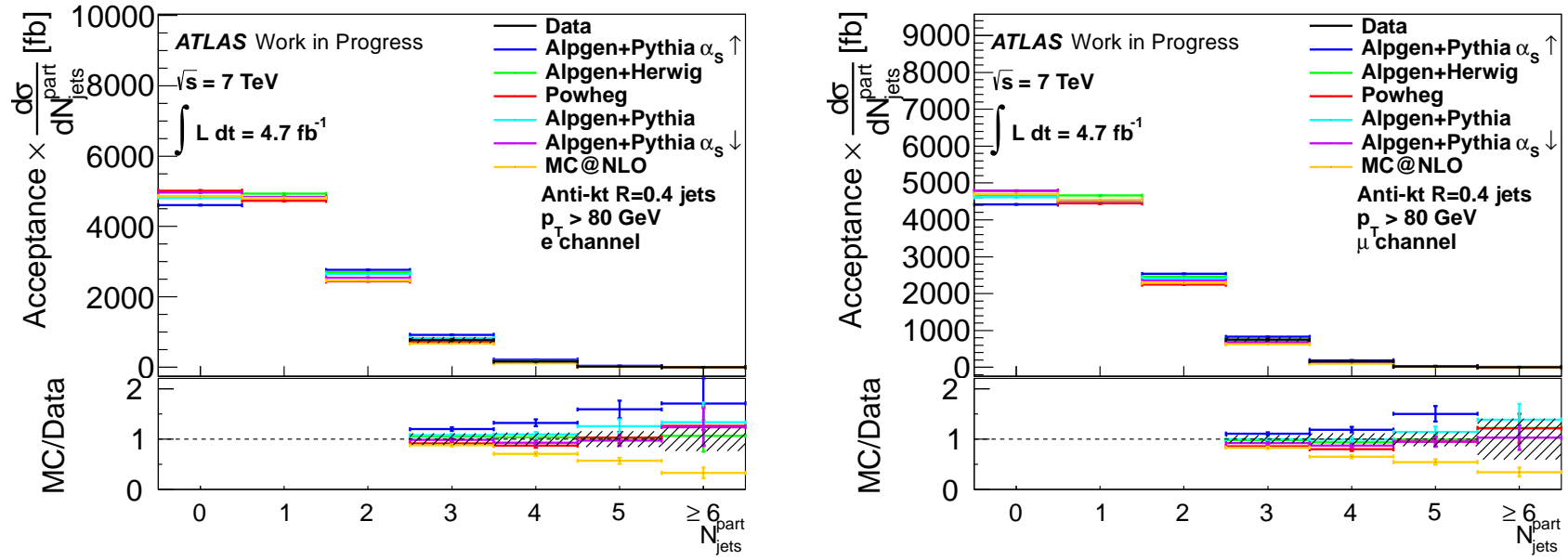


Figure 5.18: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 80 GeV.

It can be seen that the Alpgen+Pythia with the  $\alpha_S$  variation upwards overestimates data at the high jet multiplicity bins, while both Powheg and the Alpgen+Pythia with the downwards  $\alpha_S$  variation agrees very well with data in all bins for the  $p_T > 25$  GeV analysis. The nominal Alpgen+Pythia described data within the systematic uncertainty, but not as well as the version with the downwards  $\alpha_S$  variation. Only  $t\bar{t}$  with one extra parton is simulated at the MC@NLO simulation, which has a good estimate of the production cross section in the 3 and 4 jets bin. The MC@NLO simulation heavily underestimates data at large jet multiplicity, while Alpgen, which is a leading order generator, but includes up to 5 extra partons.

A complementary analysis to this one performed using the ATLAS detector is the “jet gap fraction” analysis [104], which studies the two-lepton final state of the  $t\bar{t}$  system and calculates the cross section of  $t\bar{t}$  production ( $\sigma_{t\bar{t}}$ ) and the cross section disregarding events with an extra jet produced with  $p_T > Q_0$ . The measured ratio is:

$$f(Q_0) = \frac{\sigma_{t\bar{t}}(Q_0)}{\sigma_{t\bar{t}}}. \quad (5.22)$$

This analysis also measures the effect of extra radiation in  $t\bar{t}$  decays, for the dilepton final state and it has shown that the simulation prediction has a large systematic effect, while the data uncertainty is smaller. It is interesting to compare the results of the analysis described in this document and the results for the jet gap fraction analysis. Figure 5.27 shows the value of  $f(Q_0)$  for the rapidity range  $|y| < 0.8$  and it is clear that the jet gap fraction is overestimated in MC@NLO, which means that the fraction of events in MC@NLO with extra jets is smaller than data, in agreement with the results in this document. The Alpgen+Herwig result in the jet gap fraction analysis also shows a better description of the extra radiation, although in this analysis the Alpgen+Pythia and variations are not included.

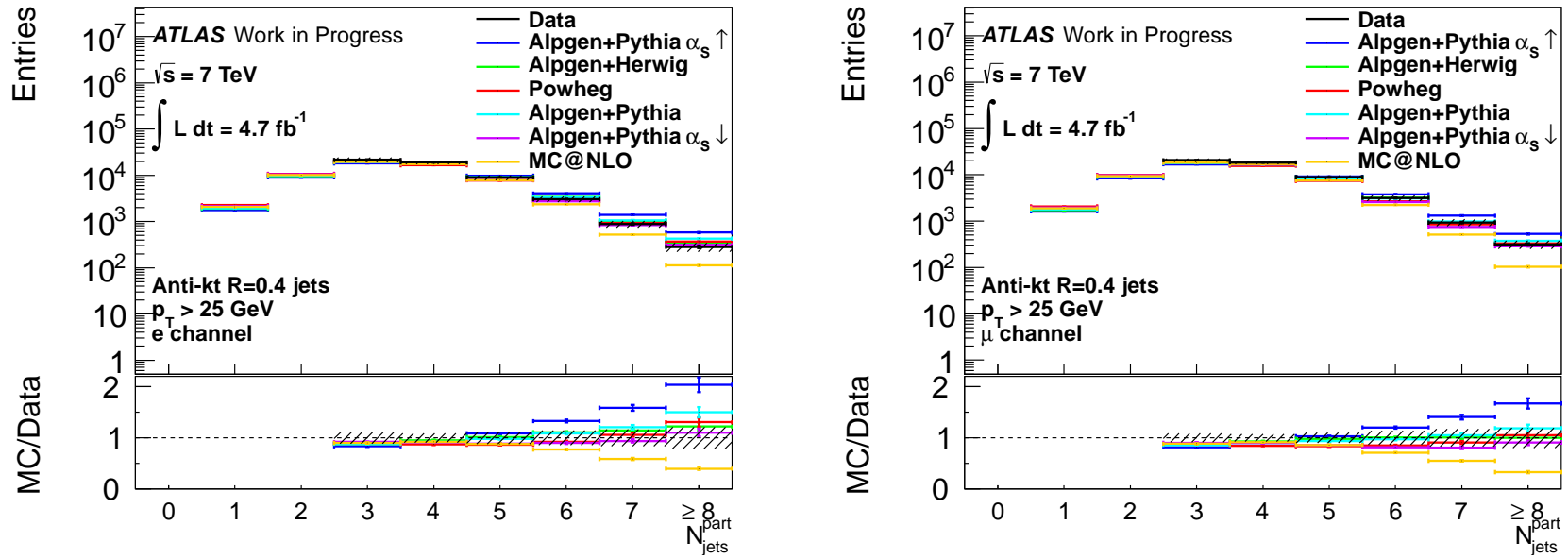


Figure 5.19: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 25 GeV.

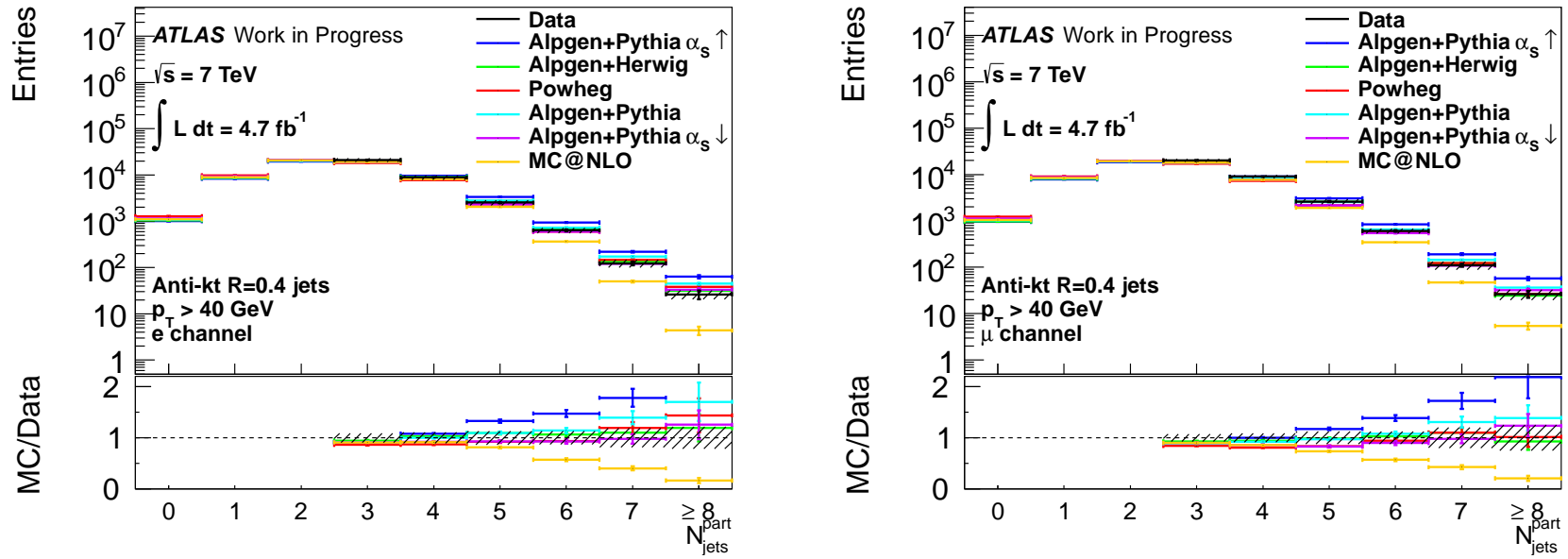


Figure 5.20: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 40 GeV.

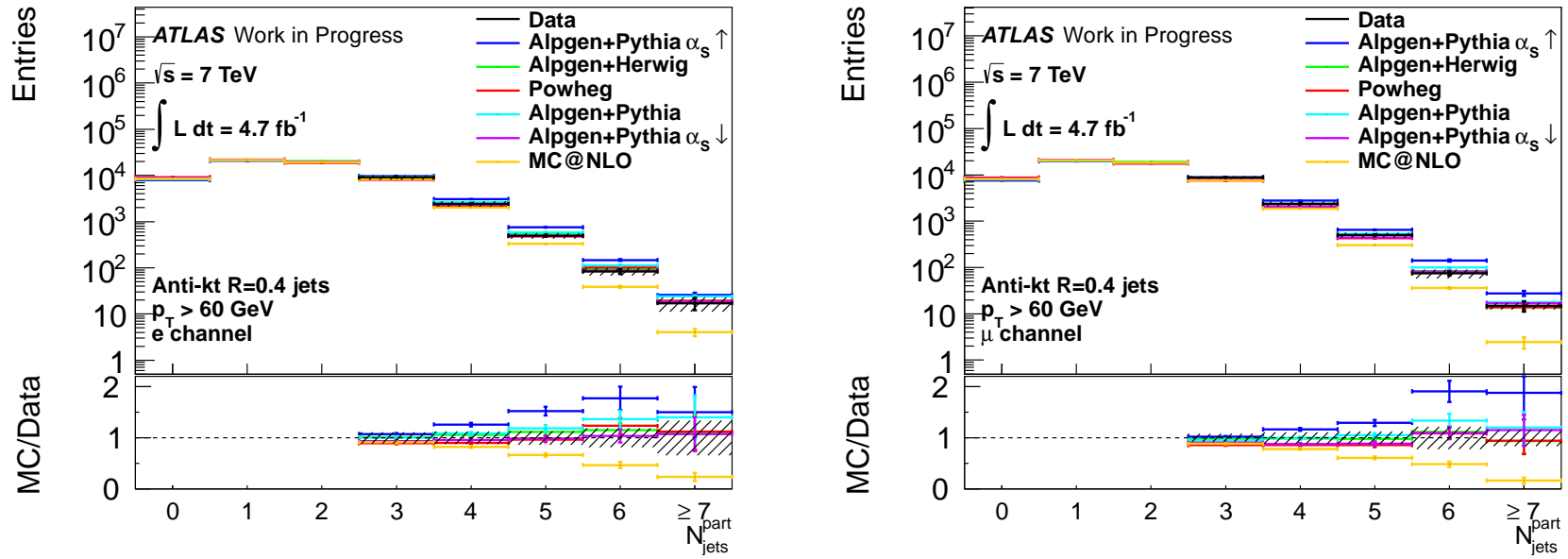


Figure 5.21: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 60 GeV.

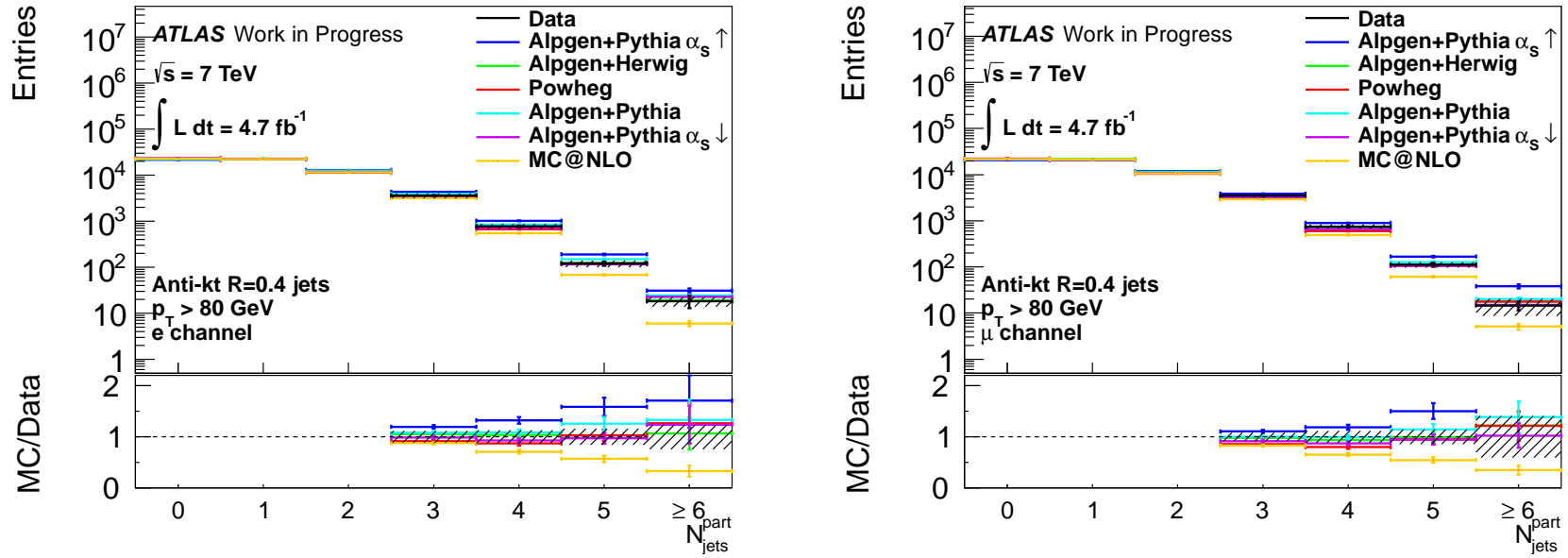


Figure 5.22: The unfolded data using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 80 GeV.

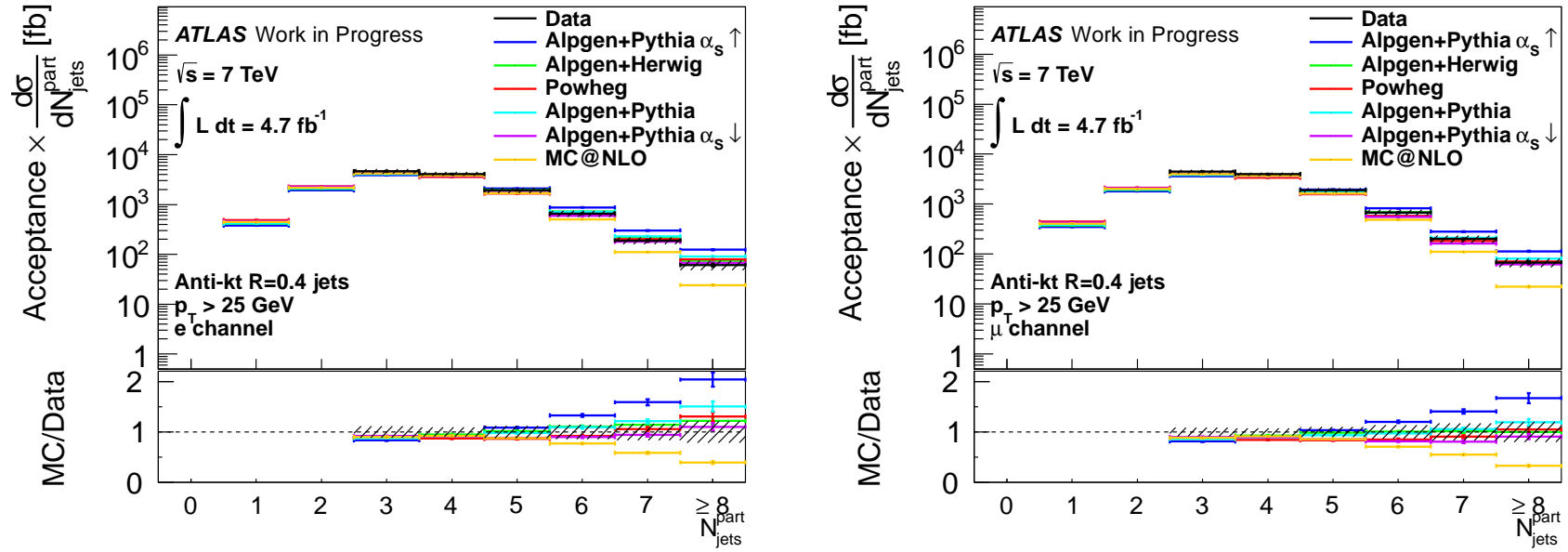


Figure 5.23: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 25 GeV.



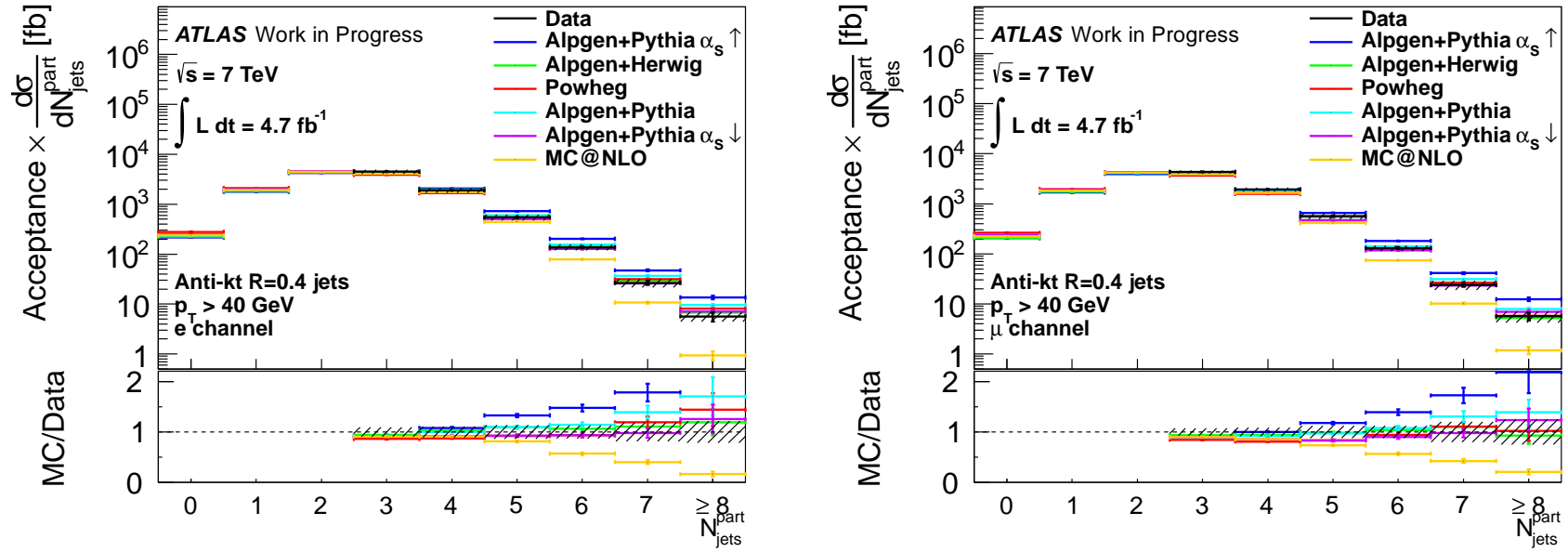


Figure 5.24: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 40 GeV.

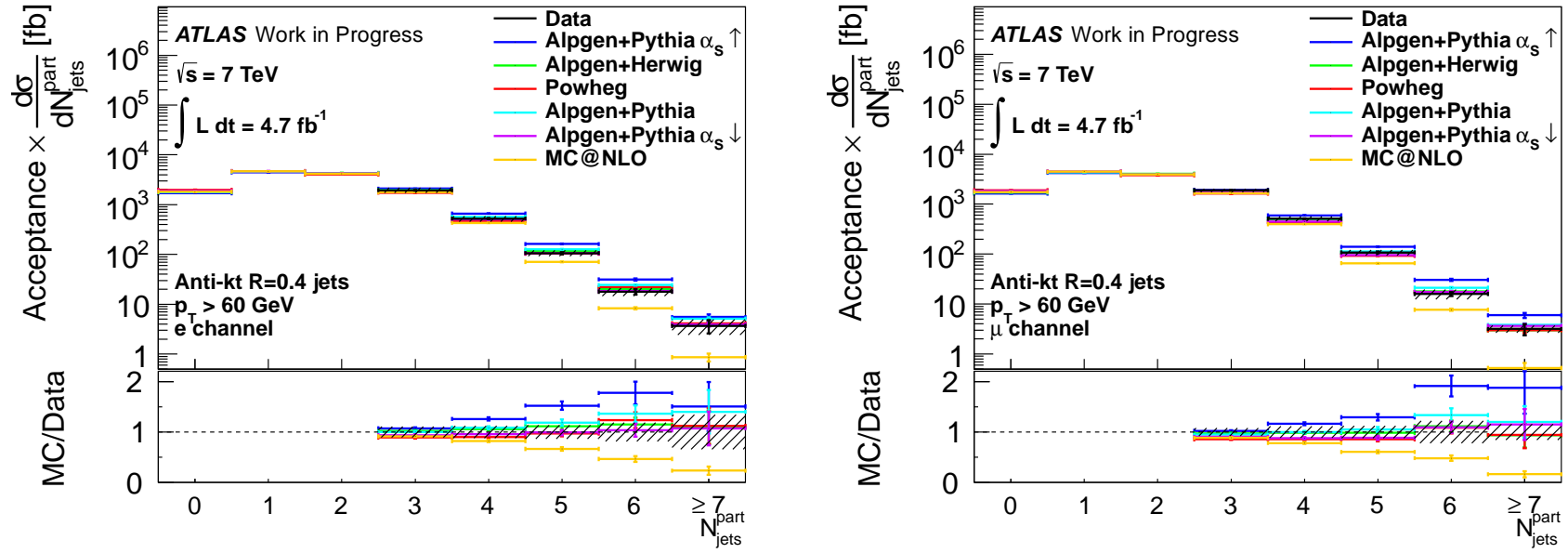


Figure 5.25: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 60 GeV.

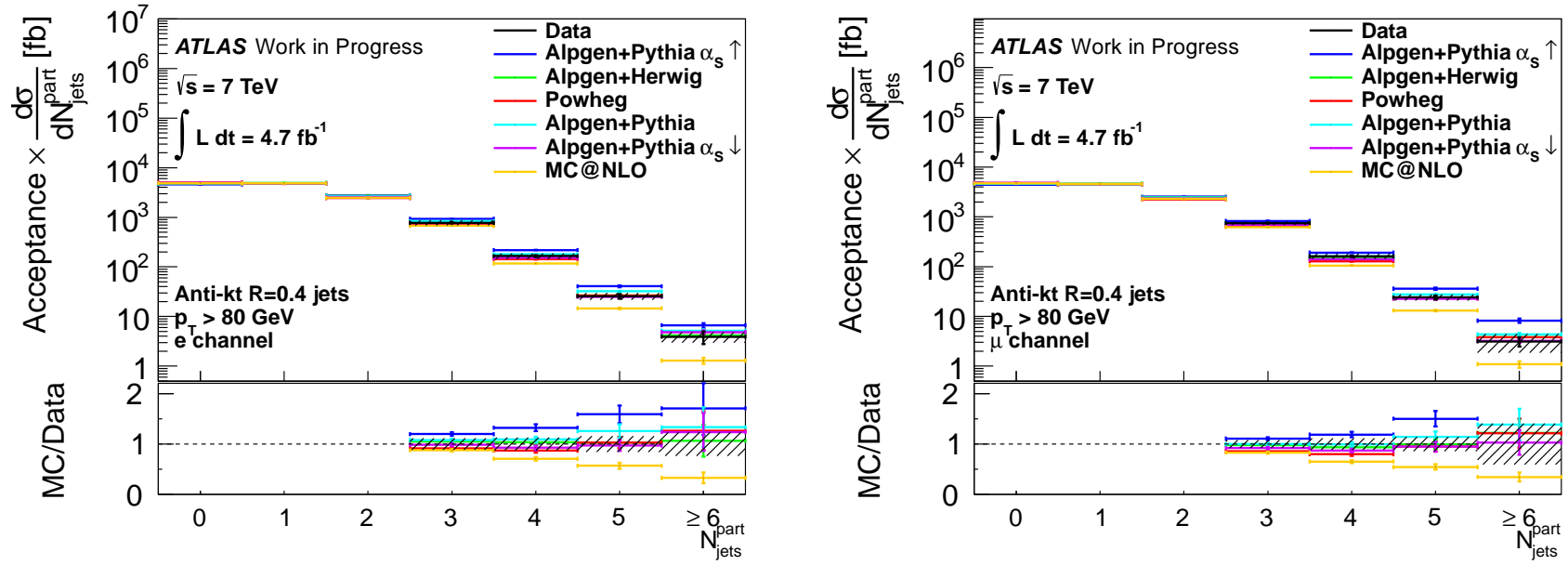
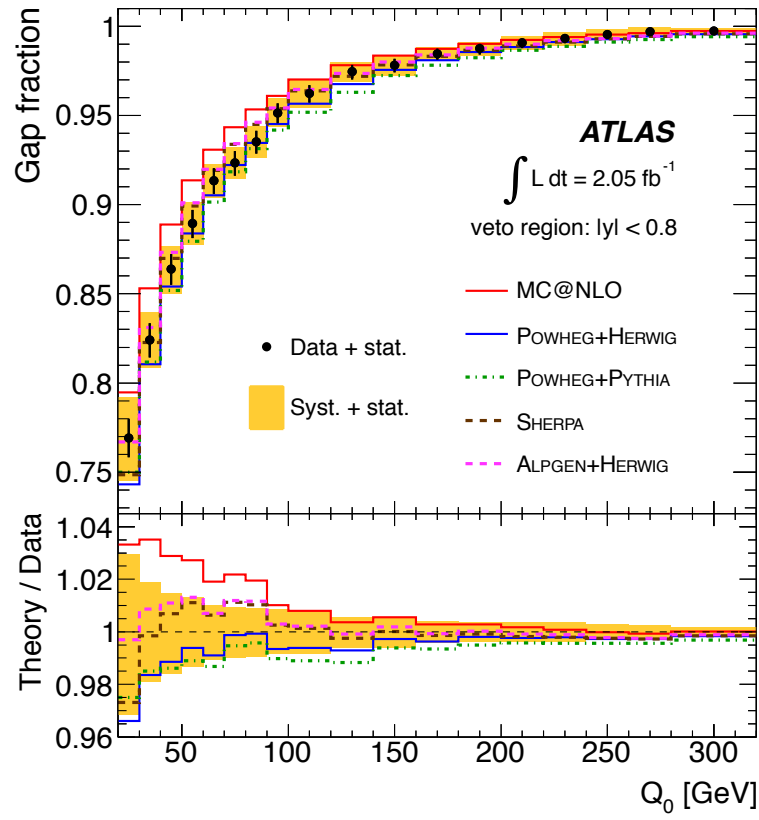


Figure 5.26: The unfolded cross section using the Alpgen  $t\bar{t}$  signal sample for corrections in logarithm scale for the Y axis. The results for the electron (left) and muon (right) channels are shown. The systematic uncertainties from reconstruction and background estimation are included. The  $p_T$  cut on the jets is 80 GeV.

Figure 5.27: Jet gap fraction for  $|y| < 0.8$ , extracted from [104].

## 5.12 Correction factors and consistency checks for selections with jet cuts at 40 GeV, 60 GeV and 80 GeV

The closure tests for selection using jet  $p_T$  cut at 40 GeV, 60 GeV and 80 GeV are shown in Figures 5.28, 5.33, 5.38. The  $1 - f'_{\text{fakes}}$  factors for the selections using jet  $p_T$  cut at 40 GeV, 60 GeV and 80 GeV are shown in the Figures 5.29, 5.34, 5.39. The  $1 - f_{\text{np3}}$  factors for the selections using jet  $p_T$  cut at 40 GeV, 60 GeV and 80 GeV are shown in the Figures 5.30, 5.35, 5.40. The  $f$  migration correction factors for the selections using jet  $p_T$  cut at 40 GeV, 60 GeV and 80 GeV are shown in the Figures 5.31, 5.36, 5.41. The  $f_{\text{reco}}$  acceptance correction factors for the selections using jet  $p_T$  cut at 40 GeV, 60 GeV and 80 GeV are shown in the Figures 5.32, 5.37, 5.42.

The systematic uncertainties for the unfolded data with the 40, 60 and 80 GeV selections are given in Tables 5.6, 5.7, 5.8, 5.9, 5.10, 5.11. All values are given as percentages of the unfolded data.

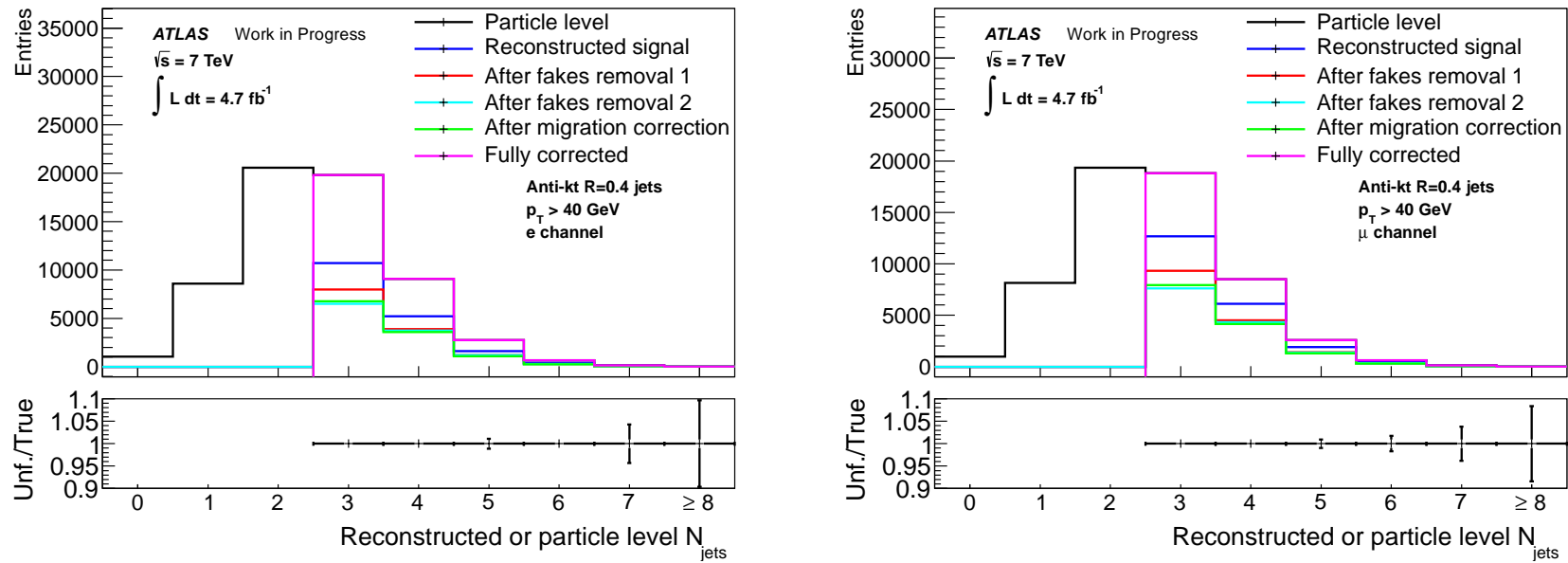


Figure 5.28: The closure test using the Alpgen  $t\bar{t}$  signal sample for input and corrections with a jet  $p_T$  cut at 40 GeV for the selection. The results for the electron (left) and muon (right) channels are shown.

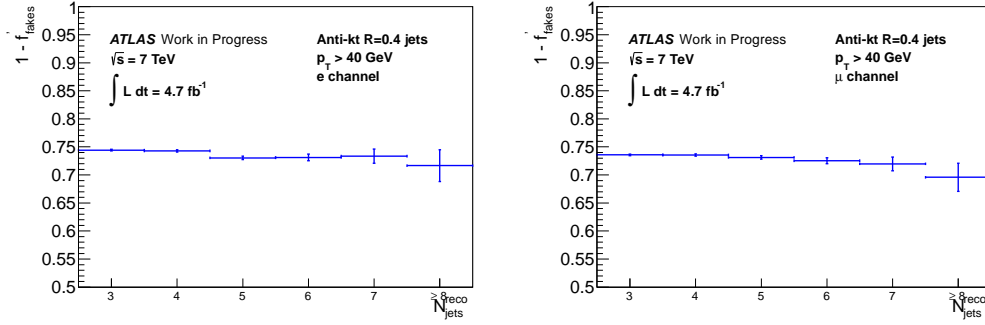


Figure 5.29: The  $1 - f'_{fakes}$  correction using the AlpGen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown.

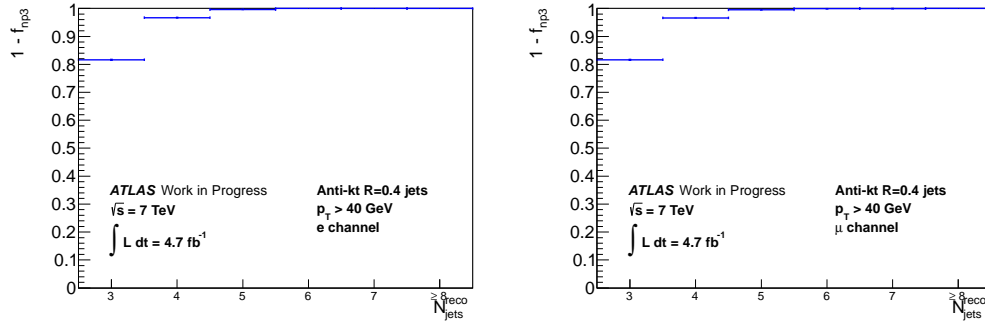


Figure 5.30: The  $1 - f_{np3}$  correction using the AlpGen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown.

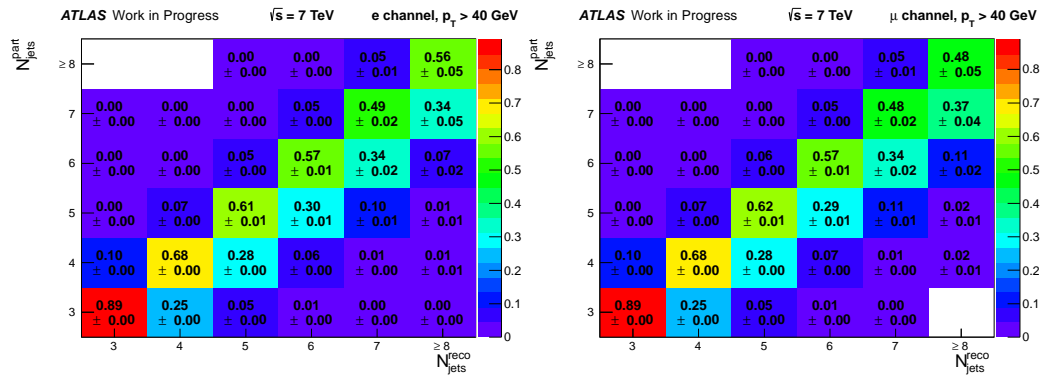


Figure 5.31: The migration matrix using the AlpGen  $t\bar{t}$  signal sample with a selection using a jet  $p_T$  cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown.

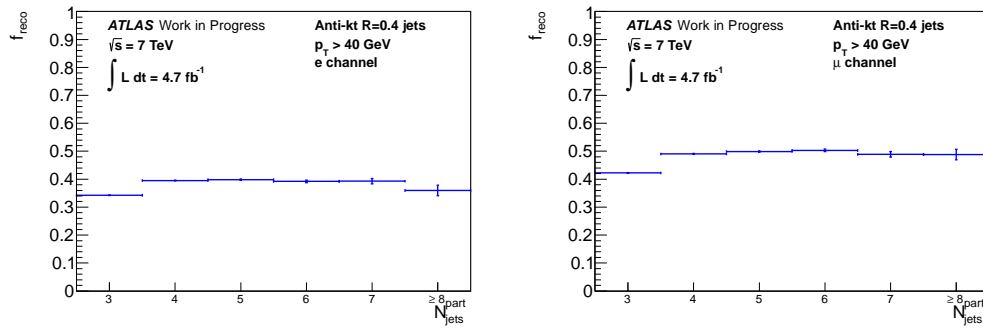


Figure 5.32: The  $f_{reco}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 40 GeV. The results for the electron (left) and muon (right) channels are shown.



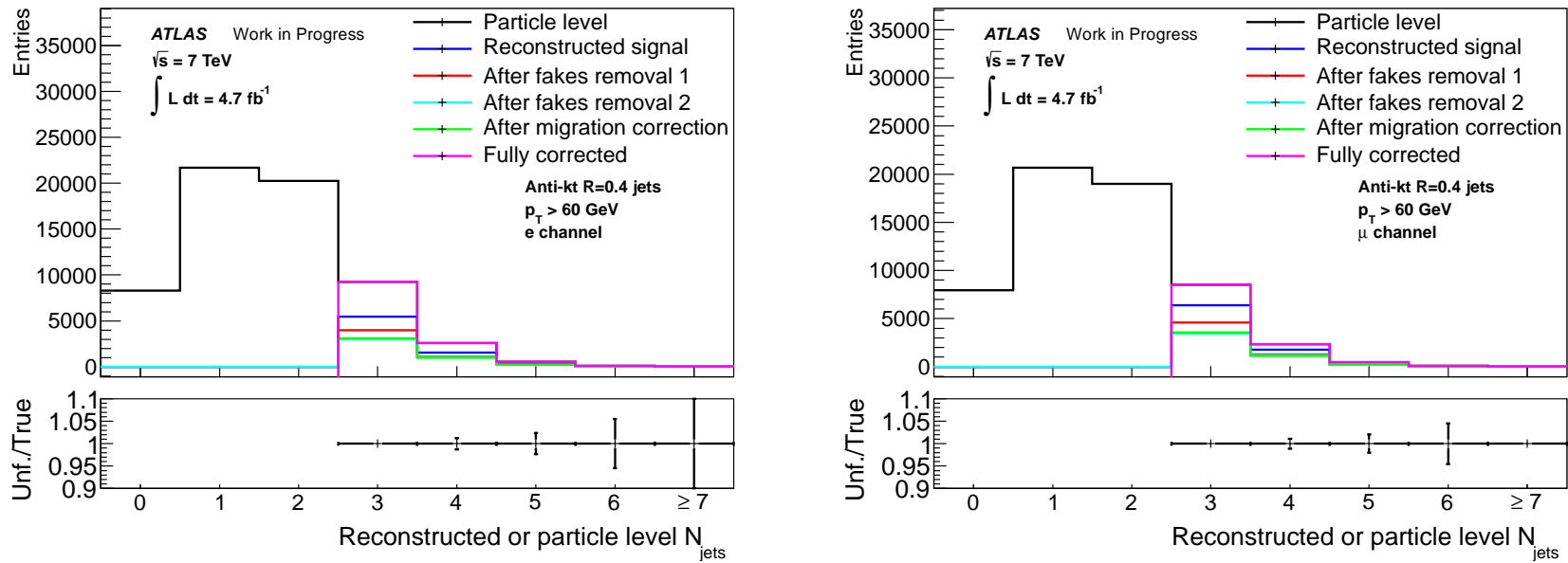


Figure 5.33: The closure test using the Alpgen  $t\bar{t}$  signal sample for input and corrections with a jet  $p_T$  cut at 60 GeV for the selection. The results for the electron (left) and muon (right) channels are shown.

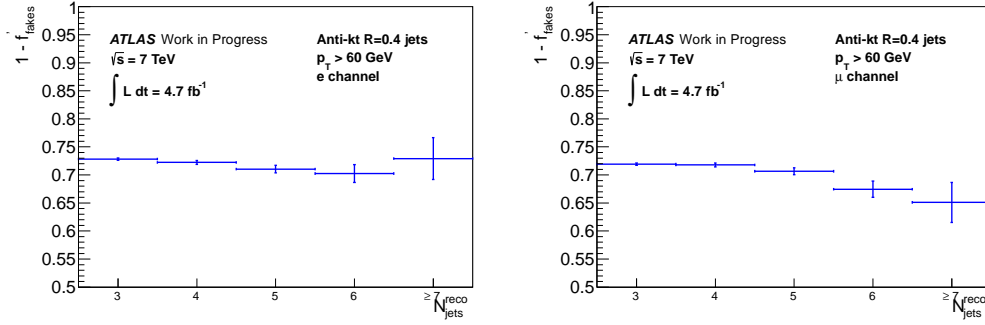


Figure 5.34: The  $1 - f'_{fakes}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown.

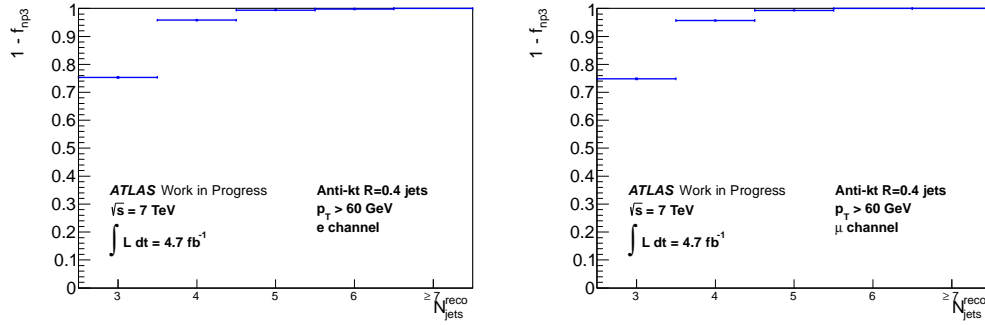


Figure 5.35: The  $1 - f_{np3}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown.

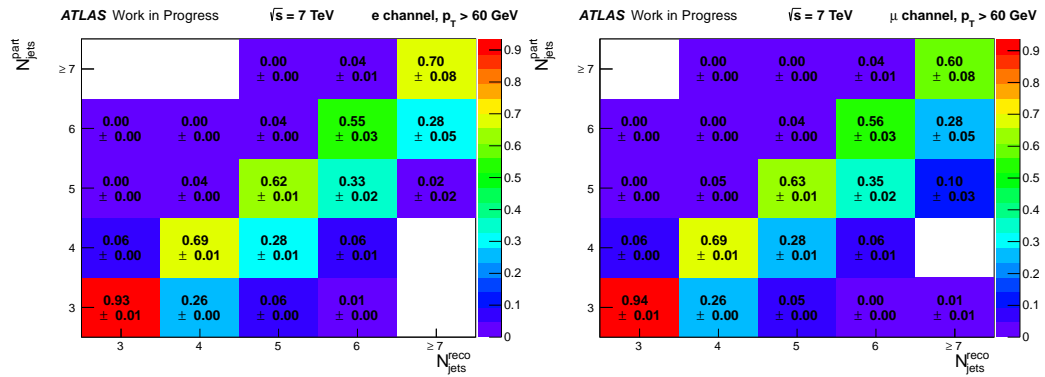


Figure 5.36: The migration matrix using the Alpgen  $t\bar{t}$  signal sample with a selection using a jet  $p_T$  cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown.

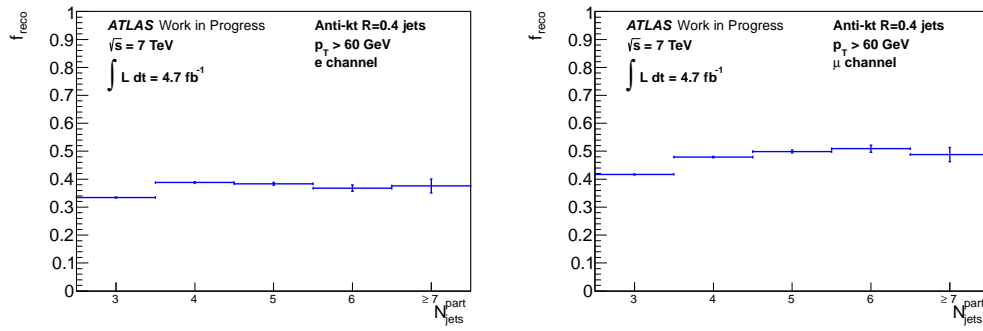


Figure 5.37: The  $f_{reco}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 60 GeV. The results for the electron (left) and muon (right) channels are shown.

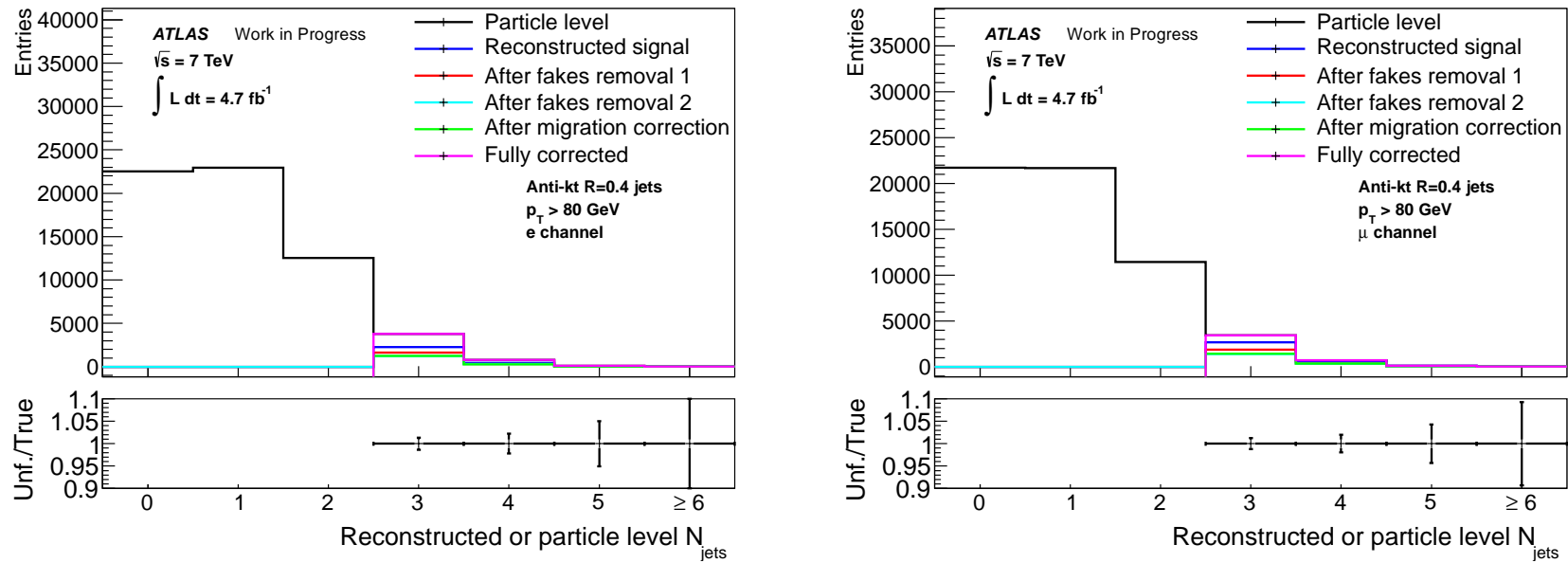


Figure 5.38: The closure test using the Alpgen  $t\bar{t}$  signal sample for input and corrections with a jet  $p_T$  cut at 80 GeV for the selection. The results for the electron (left) and muon (right) channels are shown.

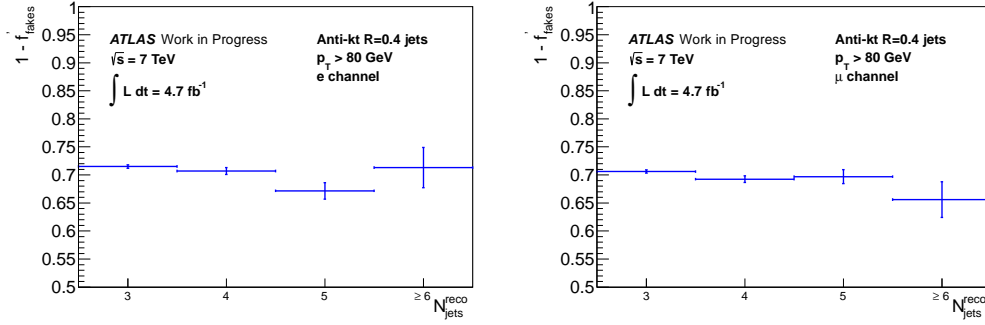


Figure 5.39: The  $1 - f'_{fakes}$  correction using the AlpGen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown.

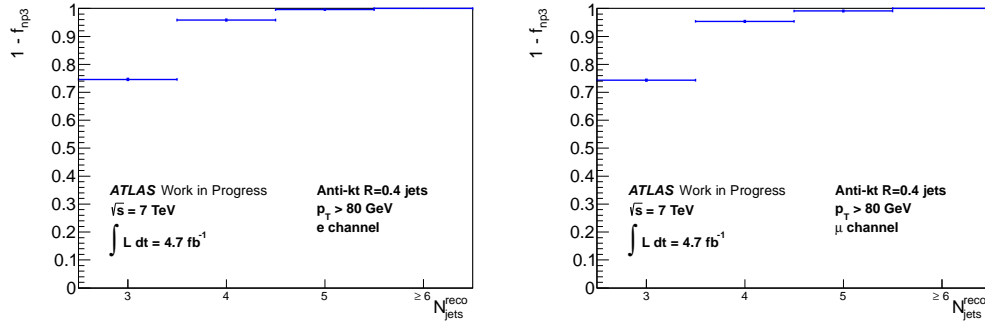


Figure 5.40: The  $1 - f_{np3}$  correction using the AlpGen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown.

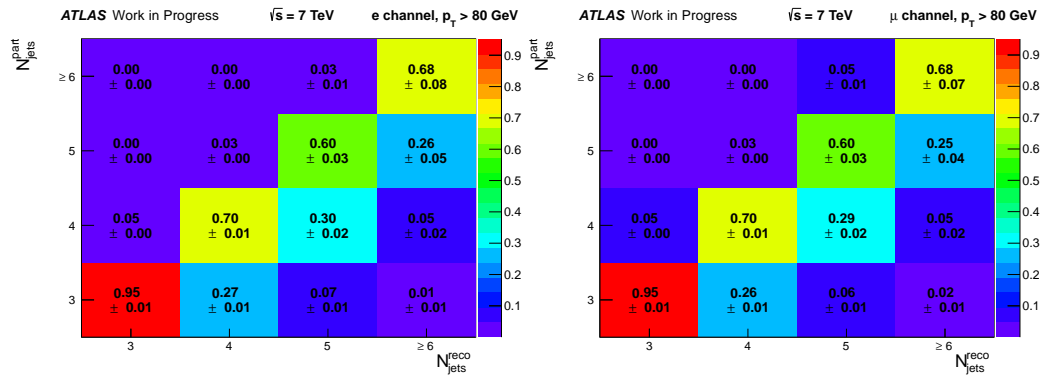


Figure 5.41: The migration matrix using the AlpGen  $t\bar{t}$  signal sample with a selection using a jet  $p_T$  cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown.

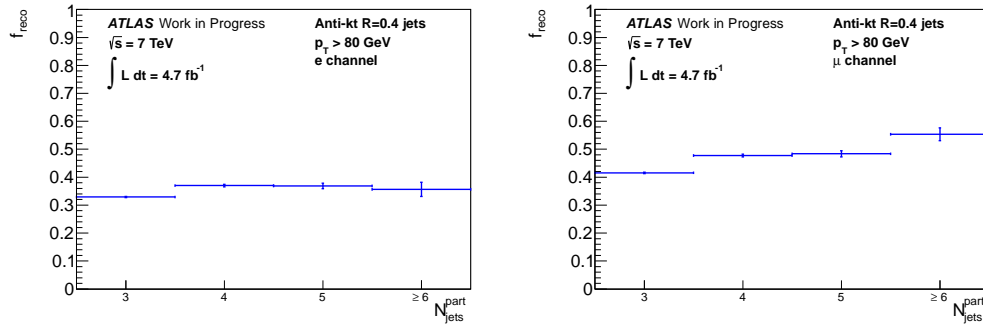


Figure 5.42: The  $f_{reco}$  correction using the Alpgen  $t\bar{t}$  signal sample with a jet  $p_T$  cut at 80 GeV. The results for the electron (left) and muon (right) channels are shown.

Table 5.6: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The  $p_T$  cut on the jets is 40 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} = 7$	$N_{\text{jets}}^{\text{part}} \geq 8$
b-tagging eff.	2.52	2.48	2.56	2.68	2.66	2.40	1.88
b-tagging c tag rate	0.14	0.12	0.15	0.22	0.25	0.14	0.55
mistag rate	0.09	0.08	0.10	0.15	0.18	0.19	0.51
Cell Out	0.03	0.01	0.05	0.05	0.08	0.29	1.11
Electron energy scale	0.02	0.01	0.03	0.01	0.07	0.10	0.14
Electron energy res.	0.20	0.20	0.20	0.22	0.22	0.23	0.36
Jet efficiency	0.02	0.01	0.03	0.09	0.04	0.05	0.03
Jet energy res.	0.36	0.29	0.12	1.17	2.29	3.40	2.51
Jet energy scale	3.16	1.94	4.28	6.99	9.83	13.55	14.86
JVF	0.80	0.73	0.85	1.00	1.17	1.34	1.51
Lepton Identification	1.20	1.19	1.23	1.26	1.28	1.30	1.30
Lepton Reconstruction	0.47	0.46	0.49	0.50	0.51	0.51	0.52
Muon mom. res. (MS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon mom. res. (ID)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Muon mom. scale	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Pile up	0.07	0.08	0.04	0.05	0.09	0.21	0.83
Luminosity	2.14	2.10	2.19	2.24	2.29	2.31	2.32
W+jets shape	0.59	0.56	0.66	0.77	0.24	0.05	0.00
W+jets norm.	1.03	1.16	0.85	0.75	0.55	0.62	0.69
W+jets bb+25%	1.10	1.29	0.83	0.60	0.42	0.36	0.16
W+jets bb+cc vs. c	0.56	0.62	0.48	0.48	0.33	0.46	0.21
W+jets c+25%	0.58	0.73	0.38	0.20	0.13	0.08	0.06
W+jets bb+cc+c vs light	1.52	1.82	1.13	0.78	0.55	0.41	0.20
DD QCD norm.	2.49	2.58	2.34	2.43	2.08	1.73	2.67
DD QCD eff. shape	0.26	0.28	0.25	0.24	0.22	0.22	0.24
DD QCD fake shape	2.18	2.28	2.02	2.06	1.92	1.20	1.98
Unfolding syst. (other MC gen.)	2.35	2.42	1.95	2.95	2.02	9.16	0.88
Unfolding syst. (same sample)	1.89	2.00	1.98	0.66	1.57	2.56	4.51
ISR/FSR	4.35	3.33	5.95	7.65	2.56	2.79	13.52
Parton shower	3.76	3.61	4.05	4.07	3.84	3.40	1.11
PDF syst.	3.16	3.10	3.10	3.59	3.93	4.22	2.61
All syst.	6.42	8.93	10.70	13.23	13.04	18.53	21.60
Stat.	0.88	1.22	1.42	2.32	4.35	8.88	21.59

Table 5.7: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The  $p_T$  cut on the jets is 40 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} = 7$	$N_{\text{jets}}^{\text{part}} \geq 8$
b-tagging eff.	2.76	2.74	2.79	2.84	2.78	2.76	2.33
b-tagging c tag rate	0.20	0.18	0.22	0.26	0.23	0.22	0.18
mistag rate	0.12	0.10	0.15	0.19	0.20	0.21	0.42
Cell Out	0.10	0.09	0.13	0.08	0.15	0.16	0.50
Electron energy scale	0.00	0.00	0.00	0.01	0.01	0.00	0.00
Electron energy res.	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Jet efficiency	0.01	0.00	0.02	0.08	0.05	0.11	0.06
Jet energy res.	0.37	0.33	0.30	0.46	1.85	3.68	3.31
Jet energy scale	3.19	1.89	4.42	6.94	9.88	13.71	14.06
JVF	0.89	0.82	0.94	1.11	1.28	1.46	1.62
Lepton Identification	0.41	0.41	0.42	0.43	0.44	0.44	0.44
Lepton Reconstruction	0.18	0.17	0.18	0.19	0.19	0.19	0.19
Muon mom. res. (MS)	0.01	0.01	0.01	0.00	0.00	0.01	0.13
Muon mom. res. (ID)	0.01	0.01	0.00	0.01	0.01	0.02	0.11
Muon mom. scale	0.21	0.22	0.21	0.17	0.12	0.15	0.07
Pile up	0.07	0.06	0.08	0.06	0.13	0.09	0.27
Luminosity	2.16	2.12	2.20	2.26	2.29	2.30	2.31
W+jets shape	0.65	0.63	0.71	0.76	0.26	0.05	0.00
W+jets norm.	0.94	1.11	0.71	0.52	0.36	0.33	0.11
W+jets bb+25%	1.05	1.26	0.79	0.52	0.29	0.32	0.12
W+jets bb+cc vs. c	0.42	0.48	0.36	0.26	0.11	0.23	0.10
W+jets c+25%	0.83	1.07	0.51	0.26	0.18	0.09	0.02
W+jets bb+cc+c vs. light	1.36	1.67	0.96	0.60	0.37	0.30	0.10
DD QCD norm.	0.68	0.77	0.55	0.47	0.40	0.43	0.34
DD QCD shape	0.55	0.59	0.50	0.45	0.41	0.41	0.37
Unfolding syst. (other MC gen.)	0.97	0.55	1.49	2.34	0.96	1.86	0.14
Unfolding syst. (same sample)	2.03	2.12	1.50	3.01	2.32	0.81	8.76
ISR/FSR	3.39	2.52	4.19	7.12	2.30	12.56	16.22
Parton shower	1.98	1.82	1.82	3.46	3.19	2.40	4.25
PDF syst.	3.28	3.27	3.19	3.40	3.92	4.79	5.99
All syst.	5.09	7.08	8.51	12.41	12.41	20.21	24.84
Stat.	0.78	1.10	1.23	1.98	3.70	7.88	16.98



Table 5.8: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The  $p_T$  cut on the jets is 60 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} \geq 7$
b-tagging eff.	2.78	2.74	2.93	2.89	3.02	2.10
b-tagging c tag rate	0.14	0.12	0.19	0.16	0.33	0.52
mistag rate	0.08	0.07	0.13	0.16	0.26	0.28
Cell Out	0.06	0.05	0.05	0.35	0.12	0.82
Electron energy scale	0.02	0.02	0.01	0.04	0.11	0.32
Electron energy res.	0.21	0.21	0.18	0.19	0.41	0.08
Jet efficiency	0.02	0.01	0.05	0.01	0.05	0.02
Jet energy res.	0.68	0.58	0.99	0.93	0.15	3.08
Jet energy scale	5.17	4.50	6.77	8.34	11.00	12.66
JVF	0.91	0.87	1.00	1.18	1.38	1.57
Lepton Identification	1.24	1.22	1.27	1.28	1.31	1.33
Lepton Reconstruction	0.50	0.50	0.52	0.52	0.54	0.55
Muon mom. res. (MS)	0.01	0.00	0.01	0.00	0.00	0.00
Muon mom. res. (ID)	0.01	0.01	0.01	0.00	0.00	0.00
Muon mom. scale	0.00	0.00	0.01	0.00	0.00	0.00
Pile up	0.13	0.14	0.05	0.24	0.04	0.59
Luminosity	2.09	2.07	2.14	2.18	2.24	2.28
W+jets shape	1.02	1.02	1.12	0.76	0.23	0.03
W+jets norm.	0.96	1.03	0.78	0.53	0.43	1.40
W+jets bb+25%	1.07	1.18	0.82	0.43	0.23	0.99
W+jets bb+cc vs. c	0.61	0.63	0.61	0.37	0.32	0.09
W+jets c+25%	0.45	0.53	0.25	0.12	0.05	0.10
W+jets bb+cc+c vs. light	1.37	1.53	0.97	0.51	0.25	1.08
DD QCD norm.	2.62	2.63	2.59	2.59	2.77	0.95
DD QCD eff. shape	0.25	0.26	0.25	0.23	0.20	0.26
DD QCD fake shape	2.18	2.19	2.18	2.17	2.20	0.69
Unfolding syst. (other MC gen.)	4.36	4.11	4.46	6.62	10.62	29.50
Unfolding syst. (same sample)	1.78	1.91	1.14	1.99	4.18	4.53
ISR/FSR	4.77	4.81	5.19	1.84	5.84	4.71
Parton shower	3.70	3.80	3.49	3.03	3.01	5.06
PDF syst.	3.28	3.20	3.41	3.87	4.44	5.09
All syst.	8.64	11.03	12.26	13.24	18.59	33.99
Stat.	1.46	1.78	2.64	5.21	11.78	29.79

Table 5.9: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The  $p_T$  cut on the jets is 60 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} = 6$	$N_{\text{jets}}^{\text{part}} \geq 7$
b-tagging eff.	3.23	3.23	3.27	3.09	2.85	3.14
b-tagging c tag rate	0.20	0.18	0.23	0.21	0.32	0.38
mistag rate	0.12	0.11	0.15	0.25	0.27	0.38
Cell Out	0.08	0.09	0.03	0.06	0.07	0.17
Electron energy scale	0.00	0.00	0.01	0.00	0.01	0.01
Electron energy res.	0.01	0.01	0.01	0.01	0.00	0.00
Jet efficiency	0.01	0.01	0.01	0.05	0.01	0.01
Jet energy res.	0.69	0.50	0.89	2.64	3.25	0.83
Jet energy scale	4.87	4.32	6.18	7.52	9.64	10.36
JVF	0.96	0.91	1.05	1.23	1.38	1.56
Lepton Identification	0.40	0.40	0.41	0.42	0.42	0.43
Lepton Reconstruction	0.18	0.18	0.19	0.19	0.19	0.20
Muon mom. res. (MS)	0.01	0.00	0.01	0.01	0.00	0.00
Muon mom. res. (ID)	0.00	0.00	0.01	0.01	0.04	0.01
Muon mom. scale	0.18	0.20	0.13	0.13	0.08	0.56
Pile up	0.04	0.05	0.00	0.11	0.20	0.29
Luminosity	2.22	2.20	2.26	2.31	2.32	2.33
W+jets shape	1.15	1.15	1.26	0.77	0.28	0.05
W+jets norm.	0.80	0.87	0.68	0.41	0.25	0.06
W+jets bb+25%	0.93	1.03	0.72	0.45	0.32	0.08
W+jets bb+cc vs. c	0.38	0.40	0.33	0.28	0.23	0.06
W+jets c+25%	0.71	0.82	0.45	0.16	0.06	0.01
W+jets bb+cc+c vs. light	1.20	1.34	0.86	0.45	0.29	0.07
DD QCD norm.	0.57	0.60	0.51	0.43	0.47	0.64
DD QCD shape	0.52	0.53	0.49	0.43	0.55	0.54
Unfolding syst. (other MC gen.)	0.56	0.40	0.58	2.73	3.10	8.06
Unfolding syst. (same sample)	0.78	0.27	1.68	4.73	6.34	0.11
ISR/FSR	2.89	1.86	6.27	2.88	16.41	2.98
Parton shower	1.42	0.93	2.59	4.04	3.69	1.72
PDF syst.	3.49	3.43	3.60	3.75	4.84	7.19
All syst.	6.14	7.62	11.07	12.26	21.83	16.00
Stat.	1.28	1.58	2.32	4.42	9.62	24.74

Table 5.10: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the electron channel. The  $p_T$  cut on the jets is 80 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} \geq 6$
b-tagging eff.	3.18	3.11	3.53	3.33	2.34
b-tagging c tag rate	0.14	0.12	0.19	0.39	0.15
mistag rate	0.11	0.09	0.14	0.21	0.21
Cell Out	0.07	0.03	0.20	0.25	0.70
Electron energy scale	0.02	0.01	0.04	0.26	0.42
Electron energy res.	0.17	0.18	0.13	0.28	0.23
Jet efficiency	0.02	0.02	0.03	0.00	0.00
Jet energy res.	1.47	1.38	1.94	0.29	8.74
Jet energy scale	5.57	5.19	6.76	8.24	11.62
JVF	0.99	0.95	1.10	1.29	1.53
Lepton Identification	1.17	1.16	1.20	1.24	1.28
Lepton Reconstruction	0.48	0.47	0.49	0.51	0.53
Muon mom. res. (MS)	0.00	0.00	0.01	0.00	0.00
Muon mom. res. (ID)	0.00	0.00	0.00	0.00	0.00
Muon mom. scale	0.01	0.01	0.00	0.00	0.00
Pile up	0.02	0.02	0.03	0.11	0.28
Luminosity	2.25	2.24	2.31	2.38	2.45
W+jets shape	1.81	1.85	1.81	1.01	0.23
W+jets norm.	1.19	1.27	0.91	0.66	0.18
W+jets bb+25%	1.31	1.45	0.83	0.25	0.02
W+jets bb+cc vs. c	0.92	1.02	0.60	0.08	0.02
W+jets c+25%	0.46	0.51	0.29	0.22	0.06
W+jets bb+cc+c vs. light	1.65	1.83	1.07	0.44	0.08
DD QCD norm.	3.01	3.04	3.01	2.21	3.73
DD QCD eff. shape	0.27	0.27	0.24	0.24	0.20
DD QCD fake shape	2.36	2.37	2.46	1.58	2.12
Unfolding syst. (other MC gen.)	4.00	3.98	3.50	8.31	1.50
Unfolding syst. (same sample)	0.59	0.46	1.10	1.22	1.51
ISR/FSR	4.53	4.32	4.87	7.92	9.88
Parton shower	4.22	4.06	5.10	2.04	14.33
PDF syst.	3.70	3.57	4.18	4.35	3.80
All syst.	9.61	11.69	13.13	15.94	23.84
Stat.	2.44	2.86	4.83	10.41	29.01

Table 5.11: Signal reconstruction systematics and unfolding bias systematics, in percentages, propagated through the unfolded distribution in the muon channel. The  $p_T$  cut on the jets is 80 GeV.

Syst.	Integral	$N_{\text{jets}}^{\text{part}} = 3$	$N_{\text{jets}}^{\text{part}} = 4$	$N_{\text{jets}}^{\text{part}} = 5$	$N_{\text{jets}}^{\text{part}} \geq 6$
b-tagging eff.	3.67	3.67	3.64	3.71	3.94
b-tagging c tag rate	0.21	0.20	0.24	0.18	0.11
mistag rate	0.16	0.13	0.25	0.40	0.22
Cell Out	0.06	0.06	0.05	0.18	0.23
Electron energy scale	0.00	0.00	0.00	0.03	0.10
Electron energy res.	0.01	0.01	0.01	0.03	0.01
Jet efficiency	0.01	0.00	0.02	0.12	0.05
Jet energy res.	0.51	0.35	1.34	0.03	1.70
Jet energy scale	4.97	4.54	6.48	7.58	8.88
JVF	1.04	1.01	1.16	1.35	1.51
Lepton Identification	0.41	0.40	0.42	0.42	0.42
Lepton Reconstruction	0.18	0.18	0.19	0.19	0.19
Muon mom. res. (MS)	0.01	0.01	0.02	0.00	0.00
Muon mom. res. (ID)	0.01	0.01	0.01	0.03	0.19
Muon mom. scale	0.16	0.17	0.15	0.03	0.00
Pile up	0.03	0.03	0.02	0.19	0.46
Luminosity	2.14	2.13	2.19	2.21	2.23
W+jets shape	1.73	1.82	1.50	0.70	0.24
W+jets norm.	0.95	1.03	0.69	0.50	0.20
W+jets bb+25%	1.22	1.33	0.82	0.54	0.22
W+jets bb+cc vs. c	0.58	0.62	0.45	0.30	0.12
W+jets c+25%	0.62	0.70	0.30	0.17	0.07
W+jets bb+cc+c vs. light	1.28	1.41	0.78	0.50	0.20
DD QCD norm.	0.57	0.59	0.48	0.59	0.50
DD QCD shape	0.50	0.51	0.46	0.54	0.55
Unfolding syst. (other MC gen.)	1.69	0.94	4.02	7.07	19.76
Unfolding syst. (same sample)	1.34	0.63	4.35	1.11	16.28
ISR/FSR	2.03	1.55	3.31	5.00	29.10
Parton shower	0.69	0.10	2.77	4.84	3.78
PDF syst.	3.69	3.66	3.76	4.14	4.34
All syst.	6.84	8.17	11.67	14.04	40.51
Stat.	2.12	2.48	4.07	9.35	21.38

## Chapter 6

# Top-antitop resonances search at $\sqrt{s} = 7$ TeV

The previous chapter has shown a measurement of the  $t\bar{t}$  cross section as a function of the jet multiplicity, which tests the Standard Model predictions related to the top quark production and the extra radiation generated in association with it.

This chapter presents an analysis which proposes a direct test of alternative models which change the  $t\bar{t}$  production due to new particles produced in the proton-proton interactions. The procedure proposed and implemented is quite general and can be used to test any model that predicts top-antitop production through non-Standard Model channels, however two benchmark models were used to set limits on the existence of the new particles. Beyond the Standard Model  $t\bar{t}$  production that does not include resonances can also be detected through this method, if it affects the observable under study.

The method used to implement this analysis consists mainly of a selection that suppresses the Standard Model backgrounds, a background estimation procedure, a reconstruction procedure for the invariant mass of the top-antitop pair system and a limit setting procedure. The final result consists of an exclusion set for the mass of the top-antitop system,  $m_{t\bar{t}}$ , in which the tested models' particles are excluded with a confidence level greater than 95%, assuming a cross section for the non-Standard Model production, given by the model. This analysis was published in [105] with a large collaboration from different researchers. The focus, in this document, will be in the author's contribution to the search.

## 6.1 Benchmark models and motivation

Although the Standard Model has many successes, there are currently open questions and it is still believed to be an effective Quantum Field Theory. For example, the effect of gravity is still not included in the Standard Model. Furthermore, no explanation for Dark Matter and Dark Energy are included in it and there are still alternate models for the electroweak symmetry breaking, besides the ones that include a scalar Higgs boson directly. There are many searches for Beyond the Standard Model physics, which test the compatibility of data and new models and they also serve to test the agreement between data and the Standard Model predictions.

Although the procedure used in this analysis is quite general, two benchmark models are tested. One benchmark model generates a top-antitop pair with high mass from the decay of a leptophobic  $Z'$ -like particle. The model is topcolor assisted technicolor (TC2) [66–68]. Another benchmark model is a Randall-Sundrum (RS) warped extra-dimensions, which includes a bulk Kaluza-Klein (KK) gluon [69, 70] that decays to a high mass top-antitop pair. The resonances might have different widths, which are related to free parameters in the model, leading (theoretically) to the resonance peaks being narrow or broad. A broader peak would be detected less easily and must have its hypothesis tested in the limit setting procedure separately.

## 6.2 Search strategy

The final state observed for the resonances in the models being probed is the same as the Standard Model  $t\bar{t}$  decays and, in this analysis as well as the one described in Chapter 5 the final state is the semileptonic decay of the  $t\bar{t}$  system. As mentioned in Section 5.2, in this channel, one of the top quarks decays into a  $b$ -quark and a  $W$ -boson, which decays into a lepton and a neutrino. The other top quark generates a final state with two quarks as a result of the  $W$ -boson decay and a  $b$ -quark. As in the case of the analysis in Chapter 5, “lepton” is used to refer to the electron or muon in the leptonic decay of one of the top quarks, including electrons or muons from the leptonic decays of the tauon. Events in which the  $t\bar{t}$  system decays semileptonically, generating a tauon which decays hadronically are regarded as background events.

The observable that is used to detect the  $t\bar{t}$  resonances in this analysis is the invariant mass of the top-antitop system ( $m_{t\bar{t}}$ ), which should have an excess above the background estimate at the resonance mass. The invariant

mass of the resonances in the benchmark models are not parameters fixed by the models, which means that a compatibility test for this excess for a range of resonance masses must be done. However, low resonance masses have already been tested in previous analyses, so this analysis will focus on the high  $m_{t\bar{t}}$  region.

If the top quarks have very high energy and hence large boost then the decay products are more collimated making it difficult to assign them to the different elements of the top decay. This situation is particularly severe in the hadronic top decay, in which three jets (a  $b$ -jet and two other jets from the  $W$ -boson decay) are merged in a single region of the calorimeter with high energy deposition. If no special treatment is given to this final state configuration, these events might be rejected, as background candidates, or  $m_{t\bar{t}}$  might be poorly estimated for them.

For the current analysis, events with large values of  $m_{t\bar{t}}$  suffer a contribution from the highly boosted top quark configuration and they are relevant to detect large invariant mass resonances, so it is important not to disregard them. With that aim, the event selection is separated to include events in two different final state configurations: the *resolved* scenario and the *boosted* scenario. In the resolved scenario, the top quark does not decay in a way that its decay products are detected very close together, so reconstructing  $m_{t\bar{t}}$  means choosing which jets are associated with the  $t\bar{t}$  system and which jets are due to extra radiation or multiple proton-proton interactions. It is not essential to identify each one of the top decay products, as long as all particles resulting from both top quark decays are selected and the  $t\bar{t}$  system four-momentum is reconstructed. The jets coming from top decay products in the resolved topology are Anti- $k_t$   $R = 0.4$  jets in this analysis, which are referred to as “small- $R$  jets”.

The boosted topology is characterised as having both top decays merged in the same region of the detector. The neutrino is not detected (except through missing transverse energy), but the lepton and the  $b$ -jet from the leptonic decay of the top might be very close together, which results in loosening the lepton isolation criteria. The  $b$ -jet from the leptonic decay of the top quark is reconstructed as a small- $R$  jet, as in the resolved scenario. The hadronic top decay in this topology should have all three quarks from the top decay in a single “large- $R$  jet”, which in this analysis is an Anti- $k_t$   $R = 1.0$  jet.

The selection strategy focuses on enforcing an orthogonalisation between the boosted events and the resolved events, since there may be an overlap between the two topologies. The large- $R$  jet in the boosted topology may be

reconstructed as one or more small- $R$  jets and the effect of the extra radiation could also increase the chances of selecting an event as if it had the resolved topology, even though it has a boosted topology. This orthogonalisation is enforced by checking if the event satisfies a boosted selection and tagging it as a boosted event if it does, regardless of whether it also satisfies the resolved selection as well. If the event does not pass the boosted selection and it does pass the resolved selection, it is tagged as a resolved event. This procedure splits the analysis in four channels, depending on whether the final state lepton is an electron or muon and on whether the event topology is boosted or resolved: *electron-channel boosted*, *muon-channel boosted*, *electron-channel resolved* and *muon-channel resolved*.

After the event has been selected and tagged in one of these four categories,  $m_{t\bar{t}}$  must be calculated from the objects available in the event. The calculation includes the neutrino and lepton four-momenta, but it also includes some of the many jets in the event, since some jets are created due to extra radiation. The missing transverse energy is used to estimate the neutrino's transverse momentum, but its  $z$ -component is not known and it can be found by applying a constraint on the  $W$ -boson invariant mass. Finally, once  $m_{t\bar{t}}$  is estimated for all channels, the spectrum with all systematic variations and contributions from all backgrounds is tested for peaks corresponding to simulation of resonance in each of the mass parameter configurations. Using the theoretical  $t\bar{t}$  production cross section given by the Standard Model and the tested model for each configuration, it can be established that the model is excluded with 95% Confidence Level for a certain resonance mass parameter range.

The data which is used to test the new models, was produced from proton-proton collisions with centre-of-mass energy of 7 TeV. The integrated luminosity for events that satisfy good detector operation quality criteria is  $4.7\text{fb}^{-1} \pm 0.2\text{fb}^{-1}$  and it was collected with the ATLAS detector in the year of 2011. It was only recorded for events with the whole detector system operational and under stable beam operations.

## 6.3 Background modelling

As the analysis' final state particles are quite similar to the  $t\bar{t} + \text{jets}$  cross section measurement discussed previously, the background sources are also quite similar. In contrast with the previous analysis,  $t\bar{t}$  is the largest background and not a signal. Furthermore, the description of the event kinematics is quite



important and the MC@NLO v4.01 simulation [57] is used to model the  $t\bar{t}$  background at next-to-leading-order. The choice of MC@NLO for the simulation of  $t\bar{t}$  events was guided by its good description of the highest transverse momentum jets' kinematics, which are important for a good estimate of the  $m_{t\bar{t}}$  observable. The description of the jet multiplicity spectrum is not excellent in MC@NLO, as was seen in Chapter 5. However, for the current analysis, only a maximum of four jets are used to calculate the  $m_{t\bar{t}}$  observable and for events with  $n_{\text{jets}}^{\text{reco}} \leq 4$ , the MC@NLO description agrees (within uncertainties) with data. The number of events in this region is much larger than the ones for high jet multiplicity and the jets' kinematic variables are very important for a good  $m_{t\bar{t}}$  estimation in the Monte Carlo simulation.

Herwig v6.520 [48, 49] is used for the parton showering and hadronisation of the  $t\bar{t}$  sample and Jimmy [63], for the modeling of multiple parton interactions. The parton distributions used must provide a good description of the high- $x$  environment, since the analysis focuses on high energy events. The CT10 [38] parton distribution functions are used and the top mass is set to 172.5 GeV. The systematic uncertainties associated with the parton showering and fragmentation are estimated by comparing the standard  $t\bar{t}$  background sample using MC@NLO with the simulation generated with Powheg [54] interfaced with Pythia [50] or Herwig [48, 49].

The second largest background is  $W$ +jets production, which is estimated through Monte Carlo simulation using Alpgen+Pythia [44, 50], but using the charge asymmetry data-driven method to estimate its normalisation (see Section 5.4). Simulation and data-driven methods are used to estimate the flavour fractions for  $W+b$ +jets,  $W+bb$ +jets,  $W+c$ +jets,  $W+cc$ +jets and  $W$ +light jets production. This is done using the same procedure as described for the jet multiplicity analysis mentioned previously, in Section 5.4. The QCD multi-jet background is estimated using the data-driven matrix method, which has already been described previously, in Section 5.5. The only difference in the procedure used is that the control region is defined differently, as described in Section 6.4.

Other backgrounds include single top production, simulated with MC@NLO, Herwig and Jimmy [48, 49, 57, 63] (as for the  $t\bar{t}$  sample) for the  $s$ - and  $Wt$ -channels; AcerMC v3.8 [43] is used with Pythia v.6.421 [50] for the parton showering and hadronisation to estimate the single top  $t$ -channel;  $Z$ +jets, simulated using Alpgen v.2.13 [44]; diboson production of  $ZZ$ ,  $WW$  and  $WZ$ , simulated using Herwig and Jimmy [48, 49, 63]. No data-driven methods are

used for these backgrounds.

## 6.4 Event selection

Events are required to satisfy a single lepton trigger (electron or muon trigger, depending on the analysis channel) for the resolved selection, or a large- $R$  jet trigger, in the boosted selection<sup>1</sup>. The single-electron trigger had a transverse momentum threshold of 20 GeV initially, but it was raised to 22 GeV later in 2011, while the single-muon trigger has a transverse momentum threshold of 18 GeV. The large- $R$  jet trigger has a transverse momentum threshold of 240 GeV. It can be seen in Figures 6.1 and 6.2 (in Section 6.8), that close to this transverse momentum value, the resolved channel has significantly less events and the boosted channel becomes relevant.

A good quality primary vertex in the event is also required, to reduce the effect of multiple proton-proton interactions. The primary vertex is identified as the vertex with highest sum of track's transverse momentum squared ( $\sum p_{T,track}^2$ ). The primary vertex is also required to have at least 5 tracks with  $p_T > 400$  MeV. Exactly one electron or muon that satisfies quality criteria must be available in the event for both resolved and boosted selections. Furthermore, the event is discarded if it contains another lepton (with the same minimum transverse momentum requirement) of the same type or a lepton of different type (that is, one electron and one muon), to reject events coming from backgrounds with two leptons and the two-lepton final state of the  $t\bar{t}$  decay from the signal.

Requirements are also made on the missing transverse energy to suppress the QCD multi-jets background. The missing transverse energy would be characterised, in the signal, as being caused by the neutrino in the semileptonic  $t\bar{t}$  decay. The missing transverse energy,  $E_T^{\text{miss}}$ , is calculated from the vector sum of calorimeter cells associated with topological clusters (see Section 3.7 for more information). As in the top-antitop jet multiplicity unfolding analysis, the transverse mass is defined as (see Section 5.3 for more information):

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \alpha)}, \quad (6.1)$$

---

<sup>1</sup>In ATLAS jargon, the large- $R$  jet trigger EF\_j240\_a10tc\_EFFS was used for the boosted selection. In the electron channel, the EF\_e20\_medium trigger was used from period B to J, the trigger EF\_e22\_medium was used for period K, and EF\_e22vh\_medium1, for periods L and M. In the muon channel, the EF\_mu18 was used from period B to I, EF\_mu18\_medium from period J to L.

in which  $p_T^\ell$  is the lepton transverse momentum,  $E_T^{\text{miss}}$  is the missing transverse energy, and  $\alpha$  is the azimuthal angle between the missing transverse energy momentum and the lepton transverse momentum. In the electron channel, the missing transverse energy is required to be larger than 30 GeV and the transverse mass,  $m_T$ , is required to be greater than 30 GeV. In the muon channel, the requirements are  $E_T^{\text{miss}} > 20$  GeV and  $E_T^{\text{miss}} + m_T > 60$  GeV. The  $m_T$  variable in this definition has an end-point at the true mother mass and this selection would emphasize the signal, which has  $m_T$  values closer to the  $W$  boson mass. Applying a cut on the sum  $E_T^{\text{miss}} + m_T$  has a better performance [105] discriminating against QCD multi-jets events, which would not have high values of  $E_T^{\text{miss}}$  and they would also not contain the  $W$ -boson leptonic decay which justifies the cut in  $m_T$ . The effect of the cut on this sum has been seen to be particularly helpful in rejecting QCD multi-jets in the muon channel.

A good quality electron is identified by the shape of its energy deposition in the Electromagnetic Calorimeter and the matching of an Inner Detector track to a Calorimeter cluster. The electron's calorimeter cluster must satisfy  $|\eta| \in [0, 1.37) \cup (1.52, 2.47]$  to exclude the transition region in  $|\eta| \in [1.37, 1.52]$ . The electron's transverse momentum must be  $p_T > 25$  GeV, to guarantee that it is above the turn-on region of the trigger efficiency. The electron's transverse energy is calculated using the cluster energy, but using the Inner Detector's track direction, to take advantage of the best resolution from each of ATLAS' subdetectors. The electron is also required to have a longitudinal distance to the primary vertex less than 2 mm, measured as the distance in the ATLAS'  $z$ -axis between the primary vertex and the point of closest approach to the electron's track, to reduce the effect of electrons coming from pile up interactions. Anti- $k_t$   $R = 0.4$  jets within  $\Delta R(\text{electron}, \text{jet}) < 0.2$  are discarded to avoid double counting of energy and electrons with  $0.2 \leq \Delta R(\text{electron}, \text{jet}) < 0.4$  are discarded to reduce the effect of non-prompt electrons from the QCD multi-jet background.

A good quality muon is identified by matching an Inner Detector track with a Muon Spectrometer track. They are required to satisfy  $|\eta| < 2.5$  and  $p_T > 25$  GeV, and to have their longitudinal distance to the primary vertex less than 2 mm, to reduce both the effect of pile up interactions and selecting muons from semi-leptonic  $B$ -hadron decays. In the resolved scenario, muons are required to have  $\Delta R(\text{muon}, \text{jet}) > 0.1$  to any Anti- $k_t$   $R = 0.4$  jet with  $p_T > 25$  GeV and  $|\eta| < 2.5$ . The muon four-momentum is calculated from the

combined fit of the Inner Detector and Muon Spectrometer tracks.

Both electrons and muons are required to satisfy an isolation requirement, to suppress non-prompt leptons. Non-prompt leptons are produced in the backgrounds, but they could be produced in the signal, in the case, for example, of a semileptonic decay of the  $B$ -hadron. Usually, the isolation requirement is enforced using the transverse energy or momentum in a fixed  $\Delta R$ -defined cone around the lepton, as it was done in the top-antitop jet multiplicity analysis (Chapter 5). However, as the top quarks become more boosted, the  $b$ -jets from the top decay become more collinear with the  $W$  boson decay products, which reduces the efficiency of the standard isolation requirement, that would reject prompt leptons close to  $b$ -jets. For this reason, a better measure of the leptons' isolation is defined using a variable cone size, called *mini-isolation*. It is defined as:

$$I_{\text{mini}}^{\ell} = \sum_{\text{tracks}} p_T^{\text{track}}, \quad \Delta R(\ell, \text{track}) < \frac{K_T}{p_T^{\ell}}. \quad (6.2)$$

In Equation 6.2,  $K_T$  is an empirical parameter chosen to be 10 GeV [105],  $p_T^{\ell}$  is the lepton's transverse momentum,  $\ell$  represents the lepton,  $p_T^{\text{track}}$  is the transverse momentum of tracks that fulfill the  $\Delta R(\ell, \text{track})$  requirement between the lepton and the track. The mini-isolation requirement on leptons is  $I_{\text{mini}}^{\ell}/p_T^{\ell} < 0.05$ , which corresponds to a 95% (98%) selection efficiency on electrons (muons).

Jets are reconstructed using the anti- $k_t$  algorithm on topological clusters of calorimeter cells. Two categories of jets are used: small- $R$  jets have  $R = 0.4$  and use the EM energy scale [87], while large- $R$  jets have  $R = 1.0$  and are locally calibrated [87, 88] (or using the Local Cluster Weighting method. See Section 3.5 for more information.). The EM energy scale for small- $R$  jets corrects the four-momentum of the jet to the expected particle four-momentum, while the large substructure of the large- $R$  jets might include many particles, requiring a different approach. Both of the jet types are also corrected using in situ techniques summarised in Section 3.5. Small- $R$  jets, in this context, are not used to refer to subjets, but to anti- $k_t$  jets reconstructed with the  $R = 0.4$  parameter.

In the boosted scenario, substructure variables [88] of the large- $R$  jet are used and the local calibration ensures a better measurement of energy distribution inside the jet. The small- $R$  jets are required to have  $p_T > 25$  GeV and  $|\eta| < 2.5$ , while large- $R$  jets should satisfy  $p_T > 350$  GeV and  $|\eta| < 2.0$ .

Furthermore, the small- $R$  jets are also required to have 75% of the scalar sum of the  $p_T$  of the jets' tracks coming from the primary vertex, among all tracks in each jet. One of the small- $R$  jets is required to be  $b$ -tagged using the MV1 tagger<sup>2</sup>. The  $b$ -tagging selection is done so that, in the resolved scenario, the algorithm has a 70%  $b$ -jet tagging efficiency in simulated  $t\bar{t}$  events and a rejection factor of 140 for light-jets for a  $p_T > 20$  GeV requirement [105]. In the boosted scenario,  $b$ -tagging small- $R$  jets with  $p_T > 25$  GeV results in a 75%  $b$ -jet selection efficiency and a light-jet rejection factor of 85 [105].

The event selection so far is identical for the resolved and boosted scenarios, but they diverge in the jet requirements that follow. In the resolved selection, each event is required to have four small- $R$  jets, or only three small- $R$  jets, if at least one of them has an invariant mass of at least 60 GeV. In the latter case, if one of the jets has a mass of at least 60 GeV, it is assumed that it contains two quarks from the hadronic  $W$  decay. In the boosted selection, however, the  $b$ -quark and the two quarks from the hadronic  $W$ -boson decay in one of the tops are expected to have merged in a single large- $R$  jet, therefore, at least one large- $R$  jet with mass greater than 100 GeV is required<sup>3</sup>.

The first splitting scale  $\sqrt{d_{12}}$ , defined in Section 3.5, is also used in the large- $R$  jet selection. The value of  $\sqrt{d_{12}}$  can be used to identify a heavy particle decay [105], which tends to be symmetric, while QCD splittings generated in the parton shower generate almost-collinear subjets with very different transverse momenta. The large- $R$  jet is required to have a  $\sqrt{d_{12}} > 40$  GeV, to reduce the contribution from QCD (see Figure 6.5).

There are a few other requirements for the top quark that decays in a lepton, neutrino and a  $b$ -jet. The small- $R$  jet (which is expected to be the  $b$ -jet from the top decay, although no  $b$ -tagging criterion is applied) must satisfy the  $p_T$ ,  $\eta$  and jet vertex fraction criteria described previously and  $\Delta R(\text{lepton, small-}R \text{ jet}) < 1.5$ . If more than one jet satisfy these selection criteria, the jet closest to the lepton, measured through the  $\Delta R$  definition, is chosen to be the  $b$ -jet from the top leptonic decay.

Another requirement in the boosted selection is that the decay products of the two tops should be well separated. A cut on the difference between the  $\phi$  coordinates of the lepton and the large- $R$  jet guarantees that the lepton is far away from what is assumed to be the hadronic top, according to

<sup>2</sup>The MV1 tagger uses information about the impact parameters, the secondary vertex, and decay topology algorithms to select  $b$ -jets [92].

<sup>3</sup>The  $p_T > 350$  GeV and  $|\eta| < 2.0$  requirements as established previously are also required.

$\Delta\phi(\text{lepton, large-}R \text{ jet}) > 2.3$ . It is also important to have the leptonic top  $b$ -jet well separated from the hadronic top and that is done by demanding that the events satisfy  $\Delta R(\text{selected leptonic } b\text{-jet, large-}R \text{ jet}) > 1.5$ , which also enforces that the two jets do not overlap.

The QCD multi-jets data-driven estimate is performed with a different control region definition, compared to the  $t\bar{t}$  + jets analysis, described previously, although the procedure used for this estimate is similar<sup>4</sup>. Compared to the standard selection, the missing transverse energy and transverse mass of the  $W$  boson requirements are inverted and, for muons, it is required that the significance of the muon tracks transverse impact parameter satisfy  $|\frac{d_0}{\sigma(d_0)}| > 4$  to enhance the heavy flavour component in the control region. In the boosted selection, at least one large- $R$  jet with  $p_T > 150$  GeV, with inverted mass and  $\sqrt{d_{12}}$  requirements is demanded, for the event to be accepted in the control region.

## 6.5 Corrections applied to simulation and data

Most of the corrections applied to simulation or data have been discussed previously in Section 5.6 and only a few differences exist between the description of the top-antitop jet multiplicity analysis and this analysis. Only the differences will be emphasized in the current section, while all other corrections mentioned in Section 5.6 are also applied in this search.

The previous analysis did not use large- $R$  jets, while this analysis does. A few studies of anti- $k_t$   $R = 1.0$  locally calibrated topological cluster jets used in this analysis with the details on their jet energy scale calibration can be seen in [88]. Consult Section 3.5 for a brief summary. The current analysis also differs from the description in Section 5.6 in the lepton isolation requirements, since a mini-isolation is used, instead of a fixed-cone variable. A scale factor was derived for this mini-isolation requirement, calculating the ratio of its efficiency in data and Monte Carlo simulation, which is used to weight each event [105].

Other corrections are kept as in the previous analysis and to avoid repetition, they are not mentioned in this chapter. Please refer to Section 5.6 for more details.

---

<sup>4</sup>Note that at this analysis, the QCD multi-jets parametrisation and related studies were not performed by the author, therefore, the procedure is mentioned for completeness.

## 6.6 Event reconstruction

The variable one tries to reconstruct to test the hypothesis that the observed data disagrees with the Standard Model is the invariant mass of the  $t\bar{t}$  system,  $m_{t\bar{t}}$ . This variable must be calculated with the information available in each event, that is, the missing transverse energy  $E_T^{\text{miss}}$ , the lepton four-momentum and the small- $R$  and large- $R$  jets' four-momenta. The procedure is also different in the boosted and resolved scenarios, but in both cases these elements are combined to retrieve a per-event estimate of  $m_{t\bar{t}}$ .

The four-momentum for the  $t\bar{t}$  system can be calculated by adding the four-vector of the quarks in the hadronic  $W$ -boson decay, the two  $b$ -jets from the top decays, the lepton and the neutrino. The  $m_{t\bar{t}}$  can be estimated through the invariant mass of the  $t\bar{t}$  system four-momentum. The main problems are finding out which jets are associated with the top decay products, and how to estimate the neutrino four-momentum, since the lepton four-momentum has been reconstructed from its track<sup>5</sup> and the cluster energy in the calorimeters. While the quarks to four-momenta association is different in the boosted and resolved selection, the neutrino treatment is the same in both scenarios.

The missing transverse energy can be used to obtain a first approximation of the neutrino  $x$  and  $y$  momentum components, since this is the only known particle in the Standard Model that would not be detected in ATLAS. In case there is more than one neutrino, the missing transverse energy would correspond to the sum of their four-momenta. The missing energy in the  $z$  axis is not measured, so another method is used to estimate the  $z$  component of the neutrino momentum. The neutrino momentum in the  $z$  direction can still be estimated, by assuming that the  $W$ -boson that decayed leptonically is on-shell, that is, that the four-momentum of the  $W$ -boson  $p_{W,\text{lep}}$  satisfies  $p_{W,\text{lep}}^i p_{W,\text{lep}}^i = m_W^2$ , up to a good approximation (where  $m_W$  is the  $W$ -boson mass [7]). This condition can be expanded as follows:

$$\begin{aligned}
 m_W^2 &= p_{W,\text{lep}}^i p_{W,\text{lep}}^i \\
 &= E_W^2 - |\vec{p}_W|^2 \\
 &= (E_\ell + E_\nu)^2 - |\vec{p}_{T,\ell} + \vec{p}_{T,\nu}|^2 - p_{z,W}^2 \\
 &= (E_\ell + \sqrt{|\vec{p}_{T,\nu}|^2 + p_{z,\nu}^2})^2 - |\vec{p}_{T,\ell} + \vec{p}_{T,\nu}|^2 - (p_{z,\ell} + p_{z,\nu})^2, \quad (6.3)
 \end{aligned}$$

---

<sup>5</sup>Or “tracks” for the muon, which has a track in the Inner Detector and in the Muon Spectrometer, that are used in a combined fit.

where  $E_W$  and  $\vec{p}_W$  are the energy and momentum of the  $W$ -boson;  $E_\ell$  and  $E_\nu$  are the energies of the lepton and neutrino;  $\vec{p}_{T,\ell}$  and  $\vec{p}_{T,\nu}$  are the transverse momenta of the lepton and neutrino;  $p_{z,W}$ ,  $p_{z,\ell}$ ,  $p_{z,\nu}$  are the  $z$ -component of the  $W$ -boson, lepton and neutrino momenta; and the neutrino is assumed to be massless.

In Equation 6.3, except for  $p_{z,\nu}$ , all other terms are known, since the four-momentum of the lepton is estimated with the Inner Detector, Calorimeter and Muon Spectrometer, and the missing transverse energy is used for  $\vec{p}_{T,\nu}$ . The equation is quadratic in  $p_{z,\nu}$ , therefore it might have two real roots, one real root or two complex roots. If two real roots exist, the one with smaller  $|p_{z,\nu}|$  is taken <sup>6</sup>, but if there are two complex roots, there is no obvious choice for  $p_{z,\nu}$ . A possibility for the complex roots in Equation 6.3 would be the resolution of the detector, which smears the measurement of the missing transverse energy. The  $x$ - and  $y$ -components of the neutrino momentum can be rotated in the transverse plane by the smallest angle until a real  $z$ -component can be calculated. This method has been implemented to reconstruct the neutrino four-momentum, according to the approximation above. The assumption that the  $W$ -boson is on-shell should not have a significant effect. The width of the  $W$ -boson [7] is  $2.085 \pm 0.042$  GeV and the resolution of the missing transverse energy [93] is fitted in  $W \rightarrow \ell\nu$  events to be  $\sigma_{E_T^{\text{miss}}} = (0.47 \text{ GeV}^{1/2}) \times \sqrt{\sum E_T}$  (where  $\sum E_T$  is the sum of transverse energy in all calorimeter topological cluster cells and it would be of the order of  $> 100$  GeV, for the minimum lepton and jet energies, which are 25 GeV).

Once an estimate for the neutrino four-momentum is found, the correct combination of jets should be found to reconstruct the  $m_{t\bar{t}}$  variable. In the boosted scenario, the hadronic top is defined as the highest  $p_T$  large- $R$  jet, which satisfies the  $\Delta R$ ,  $\Delta\phi$ ,  $\sqrt{d_{12}}$ , mass and  $p_T$  criteria described in Section 6.4. The  $b$ -jet coming from the leptonic top decay is defined as the small- $R$  jet with lowest value of  $\Delta R(\text{small-}R \text{ jet}, \text{lepton})$ , since it is expected that the  $b$ -jet and the lepton will be very close together in the boosted top-antitop environment. From these definitions, the  $t\bar{t}$  system four-momentum can be calculated by summing the leptonic top  $b$ -jet, the neutrino, the lepton and the hadronic top large- $R$  jet. The  $m_{t\bar{t}}$  variable in the boosted scenario is defined as the invariant mass of the  $t\bar{t}$  system calculated as described previously.

The resolved scenario includes many well separated small- $R$  jets and it is not trivial to associate them to the  $t\bar{t}$  system. The procedure used is to

---

<sup>6</sup>It has been show that this leads to a better resolution for the  $m_{t\bar{t}}$  reconstruction [105].



minimise a cost function  $\chi^2$ , which depends on the assignment of jets to the top decay products. That is, a value for  $\chi^2$  is calculated for all small- $R$  jet permutations and the permutation used to calculate  $m_{t\bar{t}}$  is the one that has the least value for  $\chi^2$ . If there is no small- $R$  jet with mass greater than 60 GeV, the following definition is used for a  $\chi^2$  cost function [105, 106]:

$$\begin{aligned} \chi^2 = & \left[ \frac{m_{jj} - m_W}{\sigma_W} \right]^2 + \left[ \frac{m_{jjb} - m_{jj} - m_{t_h-W}}{\sigma_{t_h-W}} \right]^2 + \left[ \frac{m_{j\ell\nu} - m_{t_\ell}}{\sigma_{t_\ell}} \right]^2 \\ & + \left[ \frac{(p_{T,jjb} - p_{T,j\ell\nu}) - (p_{T,t_h} - p_{T,t_\ell})}{\sigma_{\text{diff}p_T}} \right]^2, \end{aligned} \quad (6.4)$$

where the parameters  $m_W$ ,  $\sigma_W$ ,  $m_{t_h-W}$ ,  $\sigma_{t_h-W}$ ,  $m_{t_\ell}$ ,  $\sigma_{t_\ell}$ ,  $(p_{T,t_h} - p_{T,t_\ell})$ ,  $\sigma_{\text{diff}p_T}$  are fitted from  $t\bar{t}$  Monte Carlo simulation, comparing the quarks from the  $t\bar{t}$  decay in simulation with the reconstructed objects. The remaining terms in the equation are calculated from all permutations of jets associated to the hadronic and leptonic top quarks'  $b$ -jets and to the hadronic  $W$ -boson decay, that is,  $m_{jj}$  represents the invariant mass of two jets (which would come from the hadronic  $W$ -boson decay);  $m_{jjb}$  represents the invariant mass of three jets (which would come from the hadronic top decay);  $m_{j\ell\nu}$  represents the invariant mass of a jet, the lepton and a neutrino (which come from the leptonic top decay);  $p_{T,jjb}$  and  $p_{T,j\ell\nu}$  represent the transverse momenta of the decay products of the hadronic and leptonic top respectively.

The  $\chi^2$  function constrains on the hadronic  $W$ -boson mass, through the  $m_{jj}$  term and in the hadronic top invariant mass through the  $m_{jjb} - m_{jj}$  term, in which the  $m_{jj}$  term is subtracted to try to reduce the correlation between the hadronic  $W$ -boson mass and the hadronic top mass. The choice of the leptonic top  $b$ -jet is done by the  $m_{j\ell\nu}$  term, while the last term applies a constraint on the transverse momentum difference between the two top quarks.

If there is one or more small- $R$  jets with a mass greater than 60 GeV, an alternate definition of the cost function is used, taking into account that the minimum small- $R$  jet multiplicity required is three and to allow the heavy jet to contain the two jets from the  $W$ -boson decay or one jet from the  $W$ -boson decay and the  $b$ -jet. The cost function in this scenario is [105]:

$$\begin{aligned} \chi^2 = & \left[ \frac{m_{jj} - m_{tjj}}{\sigma_{tjj}} \right]^2 + \left[ \frac{m_{j\ell\nu} - m_{t_\ell}}{\sigma_{t_\ell}} \right]^2 \\ & + \left[ \frac{(p_{T,jjb} - p_{T,j\ell\nu}) - (p_{T,t_h} - p_{T,t_\ell})}{\sigma_{\text{diff}p_T}} \right]^2, \end{aligned} \quad (6.5)$$

where  $m_{jj}$  in the first term is the invariant mass of only two small- $R$  jets and the  $m_{tjj}$  and  $\sigma_{tjj}$  parameters are calculated in simulation [105].

Once the  $m_{t\bar{t}}$  variable is reconstructed, four independent spectra are prepared in the boosted electron channel, boosted muon channel, resolved electron channel and resolved muon channel.

## 6.7 Systematic uncertainties

There are many sources of uncertainties in the analysis, which change the signal and background estimates of the reconstructed  $m_{t\bar{t}}$ . The hypothesis testing procedure relies on these uncertainties to exclude (or not) the benchmark models, so it is very important to have a reasonable estimate of all uncertainties. Different sources of uncertainties arise in this analysis and the independent sources are added in quadrature for a final systematic uncertainty in all  $m_{t\bar{t}}$  bins and in all four analysis channels.

The  $t\bar{t}$  production cross section from simulation has an uncertainty associated with it of 11%, which is implemented as a normalisation variation in the Standard Model  $t\bar{t}$  simulation. This is a dominant normalisation uncertainty and it was calculated using approximate NNLO in QCD with Hathor 1.2 [97], using MSTW2008 90% confidence level [41] NNLO parton distribution functions sets and PDF+ $\alpha_S$  uncertainties, according to the MSTW prescription. These uncertainties are then added in quadrature to the normalisation and factorisation scale uncertainties, which are consistent with the NLO+NNLL calculation implemented in Top++ 1.0 [97, 102, 107–110].

The W+jets data-driven estimate (Section 5.4) is dominated by the statistical uncertainty. Four variations of the flavour composition are considered, increasing the flavour fractions, based on their uncertainties (including statistical uncertainties) from the data-driven method. The background uncertainty in the control region defined for this estimate is also considered when extracting the variation for the flavour fractions. The W+jets normalisation is kept constant, using the nominal data-driven estimate, while the flavour fractions are varied. A normalisation uncertainty of 60% is associated to the QCD multi-jets estimate, based on other tests that show the difference of the nominal Matrix Method used in this analysis (Section 5.5) and other methods [105].

The single top normalisation uncertainty [98–100] is calculated to be 7.7%. The Z+jets normalisation uncertainty [111] is calculated to be 48%. The diboson normalisation uncertainty is 34%, based on the parton distribution function

uncertainty and additional uncertainties associated with each extra jet. The signal and background simulation samples are normalised to the estimated luminosity in data. The luminosity in data has an estimated uncertainty of 3.9% [78]. The luminosity variation is applied to all signal samples and background samples, except multi-jets and W+jets, which are derived through data-driven methods, as described previously.

A next-to-leading-order variation on the shape of the  $t\bar{t}$  mass spectrum is also applied, by changing the renormalisation and factorisation scales by a factor of two and renormalising to the nominal  $t\bar{t}$  cross section. This variation is applied through a reweighting procedure on the  $t\bar{t}$  sample, depending on the particle-level simulation of the  $m_{t\bar{t}}$ . PDF uncertainties on all simulation samples are estimated by taking the maximum of the  $m_{t\bar{t}}$  spectra variations after reweighting the nominal samples with the CT10 [39], MSTW2008NLO [41] and NNPDF2.3 [40] uncertainty sets at 68% confidence level, according to the PDF4LHC [102] recommendation, but keeping the nominal cross section unchanged. The total PDF uncertainty is mentioned in the next section and it can be noticed that its effect is larger in the boosted selection, partly due to the uncertainty on the parton distribution functions in the high  $x$  regime.

The jet energy scale uncertainty in small- $R$  and large- $R$  jets is also a dominant uncertainty in the analysis. This uncertainty, for large- $R$  jets, includes the variation of the jet mass scale. For small- $R$  jets, besides the jet energy scale, the jet reconstruction efficiency and the jet energy resolution are also considered. The jet reconstruction efficiency is taken into account by dropping jets randomly: a pseudo-random number generator with uniform distribution is used to get pseudo-random samples between zero and one and a jet is artificially dropped in the analysis while calculating the jet reconstruction efficiency variation if this number is bigger than the estimated jet reconstruction efficiency<sup>7</sup>. A jet energy resolution variation is considered by smearing the Monte Carlo jets' transverse momentum with a pseudo-random sample of a Gaussian with mean one and variance  $\sigma_{\text{JER, data}}^2 - \sigma_{\text{JER, MC}}^2$ , where  $\sigma_{\text{JER, data}}$  is the jet energy resolution uncertainty in data and  $\sigma_{\text{JER, MC}}$  is the jet energy resolution uncertainty in Monte Carlo. Consult Section 3.5 for more details. The jet mass scale uncertainty for small- $R$  jets is not evaluated, but it is expected to have a small effect.

The  $b$ -tagging uncertainty is incorporated as a systematic variation, by

---

<sup>7</sup>Note that this is applied as a systematic variation and not as part of the nominal spectra calculation. This duplicates the effect of the jet reconstruction efficiency, estimating its effect on the spectra. The systematic variation obtained is symmetrised in the analysis.

varying the scale factors used to correct for the efficiency and rejection rates in simulation, as mentioned in Section 3.6. An extra uncertainty is added in quadrature for high transverse momentum jets with  $p_T > 200$  GeV, in which the track reconstruction is not well modelled due to the high track multiplicity environment. The jet vertex fraction scale factor correction is also varied to account for the uncertainty in its efficiency (see Section 3.5). The leptons' minimisation selection, the lepton trigger and reconstruction efficiency uncertainties are estimated using  $Z$ -boson decays to pairs of electrons or pairs of muons in data. The uncertainties in the estimation of the missing transverse energy are also considered, taking into account the effect of multiple interactions and the correction of the clusters well separated from the physics objects<sup>8</sup>. The leptons' energy scale correction and the leptons' energy resolution are also varied, as described in Sections 3.3 and 3.4.

The effect of the initial state radiation and the final state radiation in the  $t\bar{t}$  sample is also considered, using the AcerMC+Pythia [43, 50] sample and varying the Pythia parameters consistently with measurements of  $t\bar{t}$  radiation using a veto in the extra jet production, discussed in [104]. The parton shower and fragmentation uncertainties of the  $t\bar{t}$  background are computed by comparing the samples generated with Powheg+Pythia [50, 54] and Powheg+Herwig [48, 49, 54].

The higher order electroweak corrections in the  $t\bar{t}$  background were calculated in [112] and they are used to estimate its effect in the  $t\bar{t}$  normalisation uncertainty. The  $t\bar{t}$  simulation is reweighted by a parametrisation of this correction as a function of the particle-level  $m_{t\bar{t}}$ . The difference between the reweighted and nominal  $t\bar{t}$  sample is used as a systematic uncertainty associated with the higher order electroweak corrections.

The  $W$ +jets sample includes, as well as the data-driven normalisation and the flavour fraction uncertainties, a shape uncertainty associated to renormalisation and factorisation scales. The effect is parametrised as a function of the leading jet transverse momentum and the jet multiplicity in the events, which are reweighted to estimate the effect of the changes in shape.

---

<sup>8</sup>The clusters separated from the physics objects (“CellOut” term) have a special treatment in the missing transverse energy calculation with a specific calibration procedure, as mentioned in Section 3.7.

## 6.8 Data to expectation comparison

Although the goal is to calculate the  $m_{t\bar{t}}$  spectra and use it to set a limit on the benchmark models, a few checks must be made on the kinematics of the events in data, to check that the results are consistent with the Standard Model to a first approximation. It is expected that the benchmark models, or any other model, reduce to the Standard Model as an effective theory and any change in the results would be only inconsistent with the Standard Model to a small degree. Tables 6.1, 6.2, 6.3, 6.4 show the number of expected events from each Standard Model process and the number of events observed in data. The results in this study were not corrected to the particle level and they include the fiducial cuts and all selection requirements described in this chapter. The simulation is corrected by scale factors which correct differences between efficiencies and resolutions in data and simulation, as described in Section 6.5. It can be seen that there is good agreement between the total expectation values and data, within the uncertainty.

A set of checks must be done to verify that the simulation description reproduces the observables in data within the uncertainties, so that the limit setting procedure can be used to provide reliable results. Figure 6.1 shows a good agreement between the leading jet transverse momentum in data and background simulation, in the resolved selection. Figure 6.2 also shows the leading jet transverse momentum, but in the boosted scenario. The mass of the leptonically decaying top quark is reconstructed from the lepton, neutrino and the  $b$ -jet (closest jet to lepton) in the boosted scenario and it is shown in Figure 6.3. The hadronic top is reconstructed from the mass of the large- $R$  jet and it is shown in Figure 6.4, in which the mass cut has been removed only to make this plot. The top mass plots show that the corrections applied are working as expected, since the shape of the top mass peak agrees in data and simulation. The effect of the different  $E_T^{\text{miss}}$  and  $m_T$  requirements in the electron and muon channels can be seen when comparing the two plots in Figure 6.3.

Figure 6.5 shows the  $\sqrt{d_{12}}$  variable, used in the selection of the large- $R$  jets, in which the cut on this variable was removed only to make the plot. These figures show us that one can expect the simulation to describe data well, within the phase space region under analysis.

The actual spectra are in Figure 6.6 for the resolved scenario, using the  $\chi^2$  method for the  $m_{t\bar{t}}$  reconstruction and in Figure 6.7 for the boosted scenario. These estimates of the Standard Model prediction and the signal estimates are

the main ingredients used to test the hypothesis that the Beyond the Standard Model benchmark models are valid. Figure 6.8 shows the  $m_{t\bar{t}}$  spectra summed for the resolved, boosted, electron and muon channels, with one invariant mass of each benchmark model overlayed (with their production cross sections multiplied by ten).

The spectra agree well between data and Standard Model simulation, although the large  $m_{t\bar{t}}$  region is dominated by systematic and statistical uncertainties. The b-tagging efficiency and jet energy scale (and resolution) systematic variations are important uncertainties in the spectra. In the boosted scenario, the large- $R$  jet uncertainty can reach a  $\sim 20\%$  effect, being the dominant uncertainty, followed by the parton distribution function uncertainty. The large uncertainty in the high  $m_{t\bar{t}}$  bins are due to the parton distribution function contribution, which includes variations in the CT10, MSTW and NNPDF distributions (see Section 6.7). In the boosted selection, the uncertainties in the Monte Carlo simulation samples are also affected by the large statistical uncertainties.

Table 6.1: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 7$  TeV in the resolved electron channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds.

Source	Yield
$t\bar{t}$	$19652 \pm 53 \pm 3620$
single top	$1427 \pm 11 \pm 232$
W+jets	$2745 \pm 43 \pm 1037$
QCD Multi-jets	$1644 \pm 58 \pm 986$
Z+jets	$460 \pm 11 \pm 141$
Diboson	$54 \pm 1 \pm 17$
$\sum$ Backgrounds	$25984 \pm 91 \pm 4304$
Data	$26998 \pm 164$

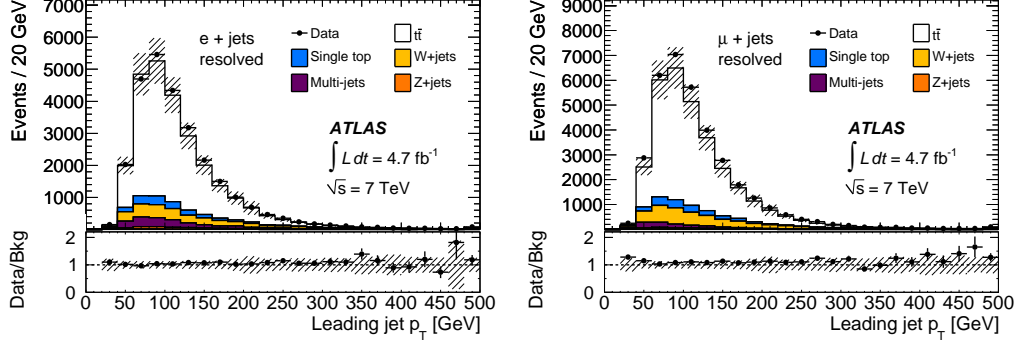


Figure 6.1: Leading jet transverse momentum in the resolved selection.

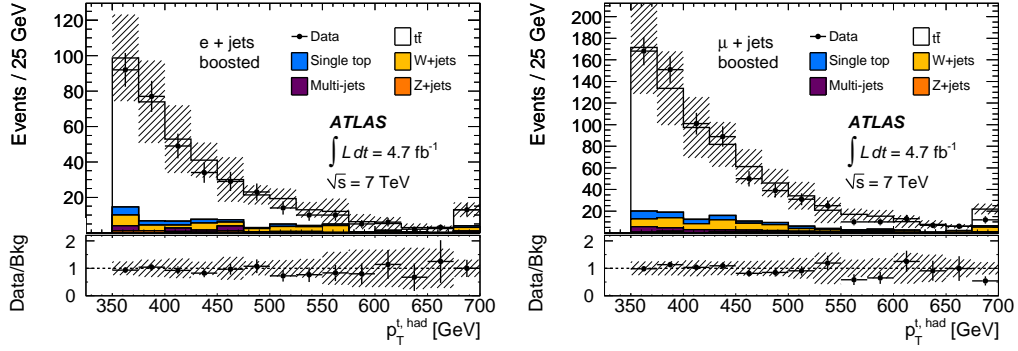


Figure 6.2: Leading jet transverse momentum in the boosted selection.

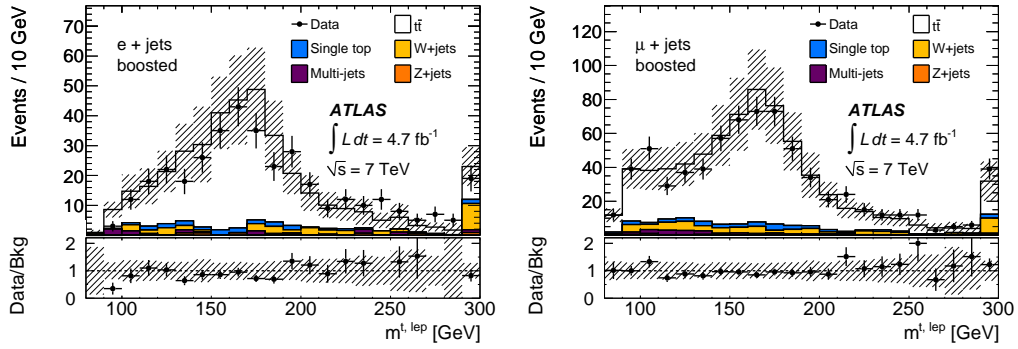


Figure 6.3: Reconstructed mass of the leptonically decaying top quark in the boosted selection.

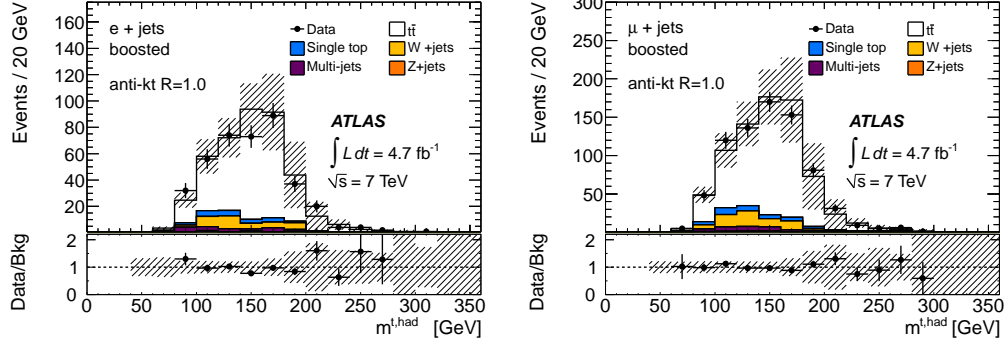


Figure 6.4: Mass of the hadronically decaying top quark in the boosted selection, reconstructed by the mass of the large- $R$  jet, with no requirement that the mass of the large- $R$  jet is greater than 100 GeV.

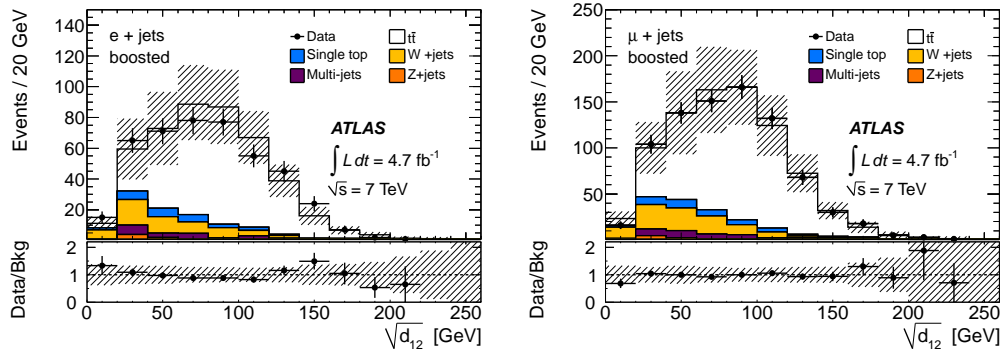
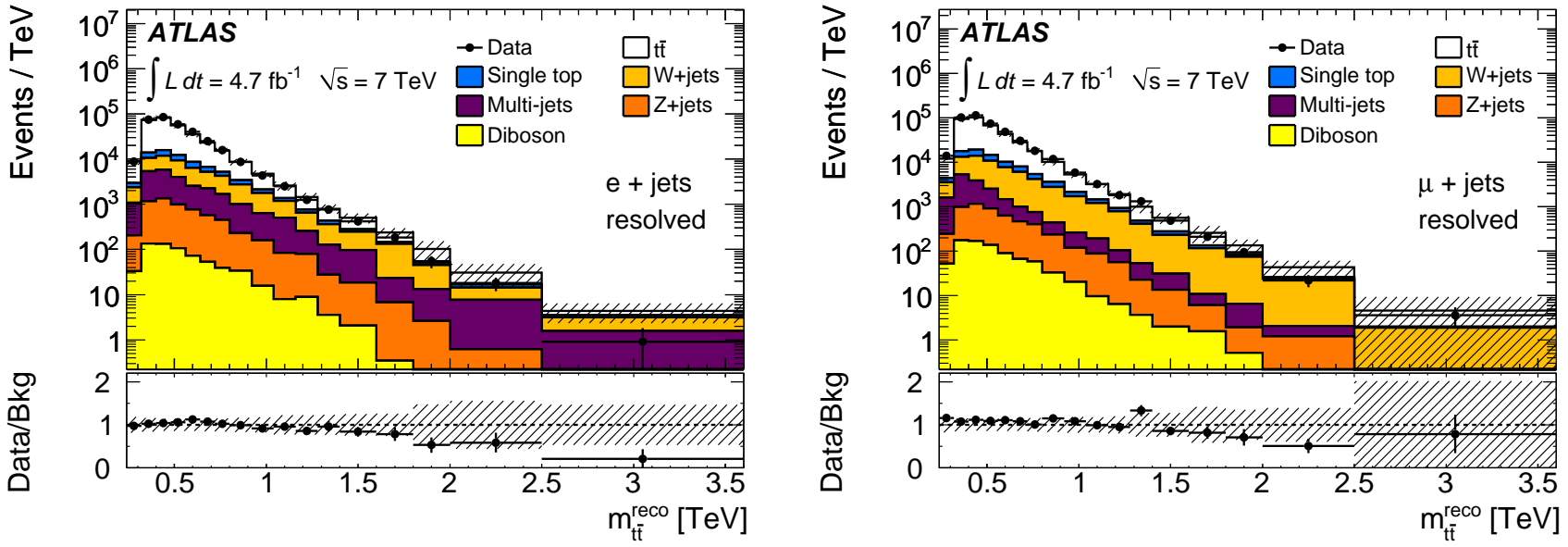
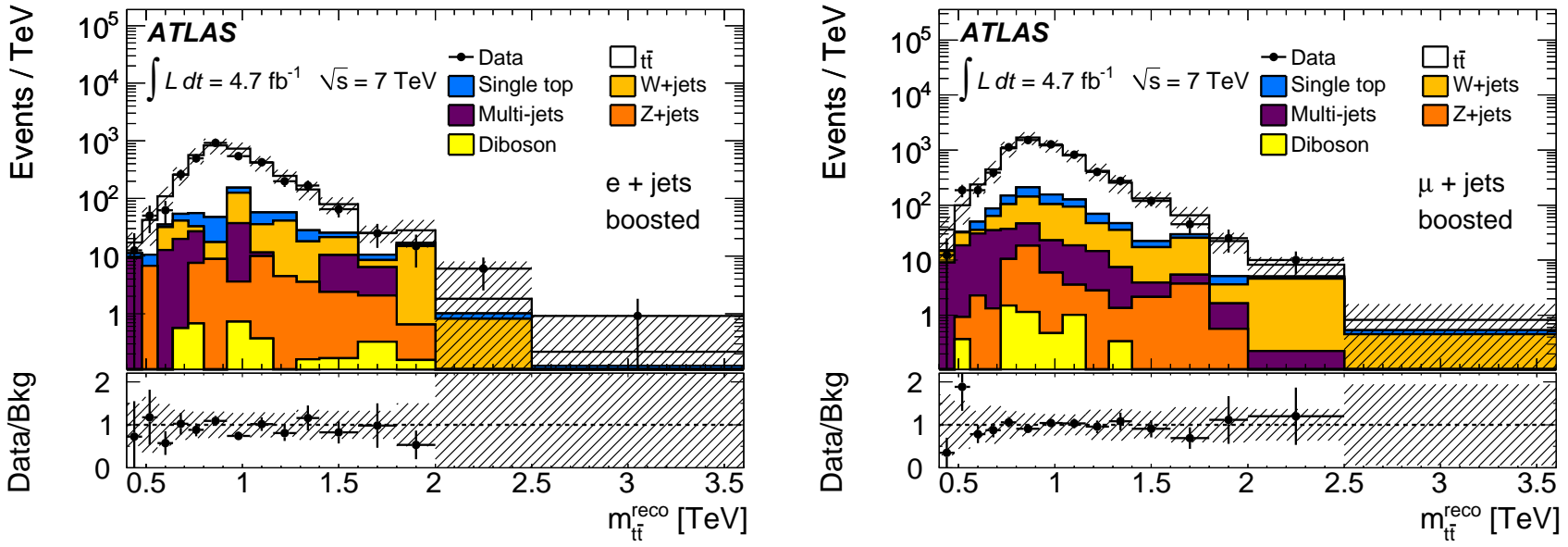


Figure 6.5: Last splitting scale for the large- $R$  jet in the boosted selection,  $\sqrt{d_{12}}$ , without the cut in this variable, in this plot.



Figure 6.6: Reconstructed invariant mass of the  $t\bar{t}$  system for selected events in the resolved scenario.

Figure 6.7: Reconstructed invariant mass of the  $t\bar{t}$  system for selected events in the boosted scenario.

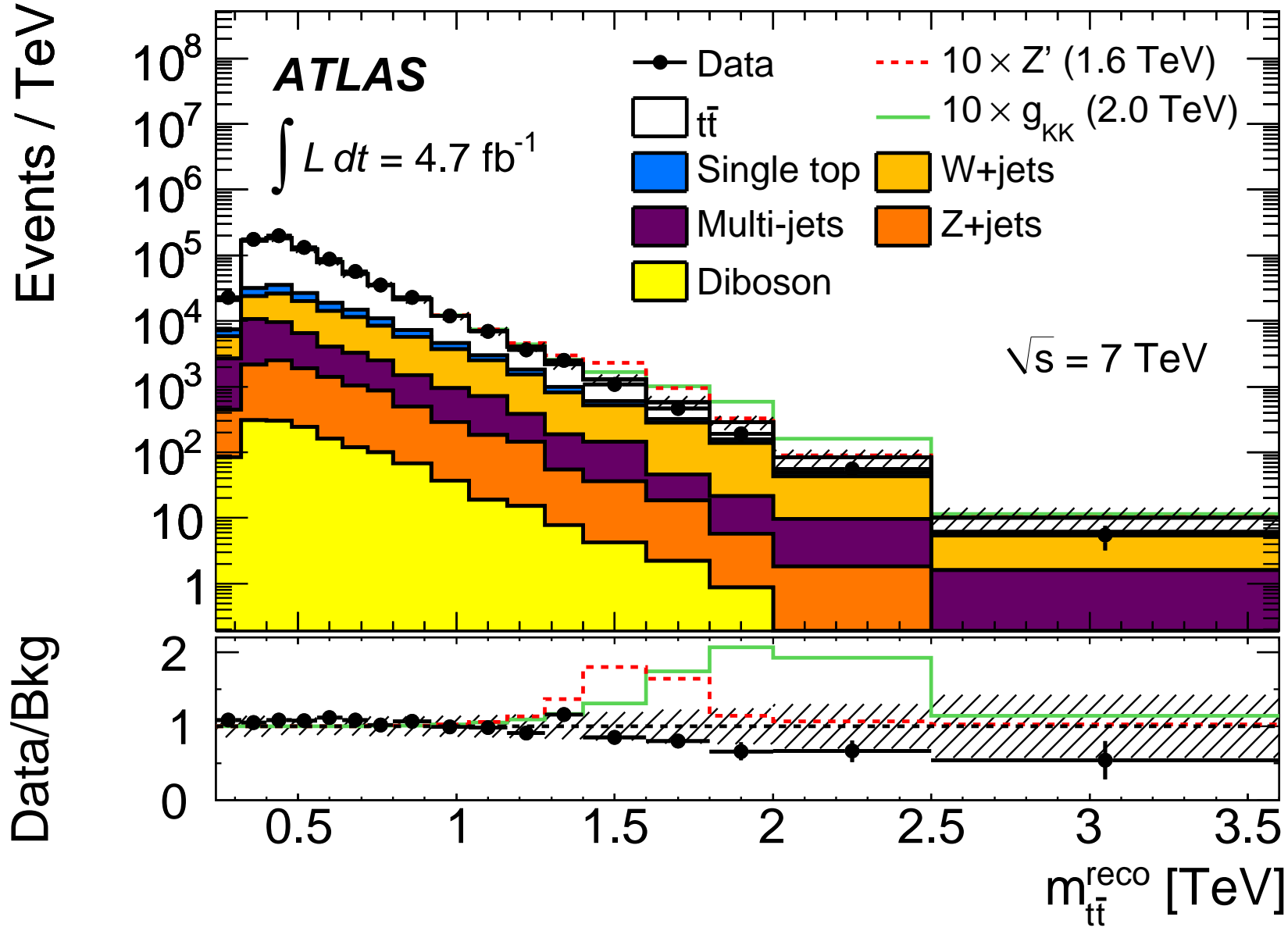


Figure 6.8: Reconstructed invariant mass of the  $t\bar{t}$  system for selected events in both resolved and boosted topologies and both electron and muon channels added in a single histogram. The  $Z'$  signal with invariant mass of 1.6 TeV and the Kaluza-Klein gluon with an invariant mass of 2.0 TeV are overlaid in this plot, with their cross section multiplied by ten to make the effect visible.

Table 6.2: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 7$  TeV in the resolved muon channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds.

Source	Yield
$t\bar{t}$	$24514 \pm 58 \pm 4418$
single top	$1802 \pm 12 \pm 295$
W+jets	$4212 \pm 54 \pm 1408$
QCD Multi-jets	$1011 \pm 12 \pm 607$
Z+jets	$383 \pm 12 \pm 110$
Diboson	$69 \pm 1 \pm 21$
$\sum$ Backgrounds	$31994 \pm 82 \pm 5257$
Data	$34956 \pm 186$

Table 6.3: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 7$  TeV in the boosted electron channel with statistical uncertainties for data and background samples, followed by the systematic uncertainty for all background samples.

Source	Yield
$t\bar{t}$	$326 \pm 6 \pm 110$
single top	$18 \pm 1 \pm 7$
W+jets	$33 \pm 4 \pm 26$
QCD Multi-jets	$7 \pm 4 \pm 4$
Z+jets	$5 \pm 1 \pm 6$
Diboson	$0 \pm 0 \pm 0$
$\sum$ Backgrounds	$392 \pm 9 \pm 125$
Data	$367 \pm 19$

Table 6.4: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 7$  TeV in the boosted muon channel with the statistical uncertainties for data and background samples, followed by the systematic uncertainty for the background samples.

Source	Yield
$t\bar{t}$	$610 \pm 9 \pm 206$
single top	$31 \pm 1 \pm 10$
W+jets	$57 \pm 5 \pm 28$
QCD Multi-jets	$19 \pm 1 \pm 11$
Z+jets	$5 \pm 1 \pm 4$
Diboson	$0 \pm 0 \pm 0$
$\sum$ Backgrounds	$725 \pm 10 \pm 217$
Data	$712 \pm 26$

Table 6.5: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the resolved electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	3.95
b-tagging c-tag rate	1.06
mistag rate	0.95
Electron id./rec.	1.45
Muon id./rec.	0.00
DD QCD norm.	3.79
Large- $R$ JES/JMS	0.25
Large- $R$ trigger	0.00
Jet energy scale	8.62
Jet efficiency	0.15
Jet energy resolution	1.57
JVF	1.43
W+jets bb+cc+c versus light	1.77
W+jets bb+25%	1.26
W+jets c+25%	0.62
W+jets norm.	1.35
Luminosity	3.24
ISR/FSR	0.42
Parton shower	0.00
$t\bar{t}$ cross section norm.	8.09
$t\bar{t}$ higher order QCD corr.	2.92
$t\bar{t}$ electroweak Sudakov corr.	1.95
PDF syst.	5.23
All syst.	17.00
Stat. MC	0.35
Stat. Data	0.63

## 6.9 Limit setting and summary

With an estimate of the signals and background  $m_{t\bar{t}}$  distributions and their systematic uncertainties, and the data  $m_{t\bar{t}}$  spectra as well, the hypothesis that the data agrees with the signal and background hypothesis can be tested using statistical methods. It is worth mentioning that, as discussed previously, the technique used so far is quite general and could be applied to test the validity of other models that include a decay to a top-antitop pair.

Although the limit setting procedure was not performed by the author,

the results obtained are quoted here for completeness. The  $m_{t\bar{t}}$  spectra shown in the previous section, with all systematic uncertainties is used to test the hypothesis of the validity of the benchmark models, using the BumpHunter tool [113]. This tool tests the hypothesis that the data disagrees with signal plus background. The systematic uncertainties are taken into account through a set of pseudo-experiments which allow the Standard Model prediction to float as within the error bands. No significant deviation from the Standard Model prediction is observed.

A Bayesian limit setting procedure developed in [114] is implemented to set the probability that the benchmark models are excluded for a certain parameter configuration in these models. The free parameters of the benchmark models are the boson's mass, which means that this procedure, which sets a Confidence Level for the signal plus background hypothesis, is repeated for different  $t\bar{t}$  resonance masses. Upper limits for the signal's cross section are set for different masses, using a uniform prior. The upper cross section limits for the benchmark models are shown in Figures 6.9 and 6.10. With these results, a  $Z'$  boson with mass between 0.5 TeV and 1.74 TeV is excluded with 95% Confidence Level. A Kaluza-Klein gluon with mass between 0.7 TeV and 2.07 TeV is excluded as well with 95% Confidence Level [105].

The results shown in this chapter include a  $t\bar{t}$  system reconstruction for events enriched in  $t\bar{t}$  decays, with a full estimate of the systematic uncertainties in the Standard Model and the signal samples. A comparison of the background with data shows no significant deviation from the Standard Model with an integrated luminosity of  $4.7 \text{ fb}^{-1}$  ATLAS data, which was collected at a center-of-mass energy of  $\sqrt{s} = 7 \text{ TeV}$ . This analysis was initially made public as an ATLAS conference note in [106] and, after a few more studies, it was published as a paper in [105].

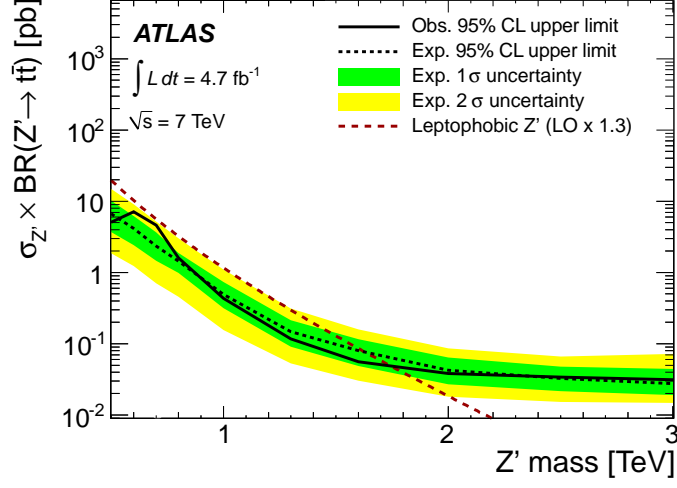


Figure 6.9: Observed and expected upper cross section times branching ratio limit for a narrow  $Z'$  resonance. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a  $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [105].

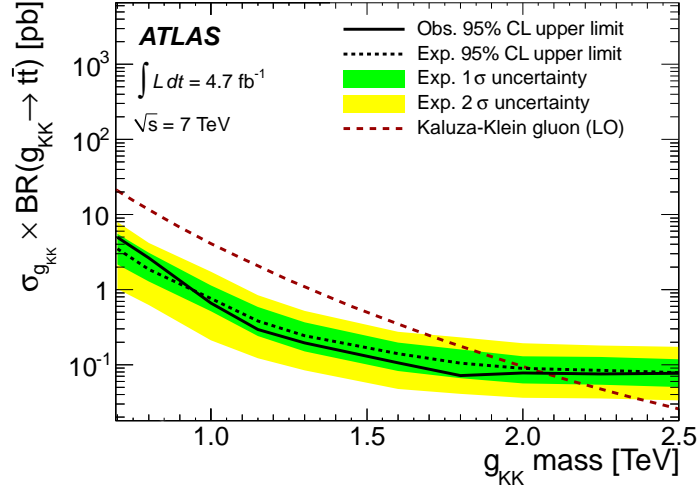


Figure 6.10: Observed and expected upper cross section times branching ratio limit for a Kaluza-Klein gluon. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a  $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [105].

Table 6.6: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the resolved muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	4.02
b-tagging c-tag rate	1.34
mistag rate	1.01
Electron id./rec.	0.00
Muon id./rec.	1.56
DD QCD norm.	1.89
Large- $R$ JES/JMS	0.24
Large- $R$ trigger	0.01
Jet energy scale	8.31
Jet efficiency	0.26
Jet energy resolution	1.26
JVF	1.50
W+jets bb+cc+c versus light	1.75
W+jets bb+25%	1.37
W+jets c+25%	1.04
W+jets norm.	1.27
Luminosity	3.26
ISR/FSR	0.48
Parton shower	0.13
$t\bar{t}$ cross section norm.	8.20
$t\bar{t}$ higher order QCD corr.	3.00
$t\bar{t}$ electroweak Sudakov corr.	1.98
PDF syst.	4.25
All syst.	16.55
Stat. MC	0.26
Stat. Data	0.58



Table 6.7: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 7$  TeV, in the boosted electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	7.57
b-tagging c-tag rate	0.90
mistag rate	0.68
Electron id./rec.	1.45
Muon id./rec.	0.00
DD QCD norm.	2.32
Large- $R$ JES/JMS	20.78
Large- $R$ trigger	0.40
Jet energy scale	5.82
Jet efficiency	0.75
Jet energy resolution	3.89
JVF	1.84
W+jets bb+cc+c versus light	1.27
W+jets bb+25%	1.04
W+jets c+25%	0.17
W+jets norm.	1.63
Luminosity	3.48
ISR/FSR	0.57
Parton shower	6.56
$t\bar{t}$ cross section norm.	8.91
$t\bar{t}$ higher order QCD corr.	9.38
$t\bar{t}$ electroweak Sudakov corr.	4.22
PDF syst.	10.44
All syst.	32.06
Stat. MC	2.37
Stat. Data	4.88

Table 6.8: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the boosted muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	6.75
b-tagging c-tag rate	0.38
mistag rate	0.95
Electron id./rec.	0.00
Muon id./rec.	0.87
DD QCD norm.	1.63
Large- $R$ JES/JMS	19.73
Large- $R$ trigger	0.67
Jet energy scale	4.04
Jet efficiency	0.86
Jet energy resolution	2.56
JVF	1.93
W+jets bb+cc+c versus light	1.12
W+jets bb+25%	1.16
W+jets c+25%	0.17
W+jets Norm.	1.41
Luminosity	3.49
ISR/FSR	1.10
Parton shower	7.90
$t\bar{t}$ cross section norm.	9.01
$t\bar{t}$ higher order QCD corr.	9.59
$t\bar{t}$ electroweak Sudakov corr.	4.31
PDF syst.	5.65
All syst.	30.04
Stat. MC	1.51
Stat. Data	3.68

# Chapter 7

## Top-antitop resonances search at $\sqrt{s} = 8$ TeV

The results in the analysis detailed in the previous chapter have been extended to the  $\sqrt{s} = 8$  TeV ATLAS data. The experience with the previous results led the group to a set of improvements which were implemented in the  $t\bar{t}$  mass spectra for the  $\sqrt{s} = 8$  TeV data, collected in 2012. Since most of the analysis procedure remains unchanged, only the small differences and the new results are quoted in this chapter. The reader can find more information about the whole analysis in Chapter 6. While in the previous analyses, the QCD multi-jets background estimate was done by other researchers, for the search described in this chapter, with 2012 data, this background was estimated by the author and a more detailed account of the method used is given. This analysis' result has been made public as an ATLAS conference note in [115].

### 7.1 Differences with respect to the $\sqrt{s} = 7$ TeV analysis

A few improvements were made to the analysis, which is fully described in [115]. The search strategy is the same as the one described in Chapter 6, separating and orthogonalising the resolved and boosted selections in the final state, resulting in four analysis channels (two for electrons and two for muons). For more details on the analysis strategy, consult Section 6.2. Section 6.3 describes the backgrounds and Section 6.4, describes the selection algorithm. Consult Sections 6.5 and 5.6 for the correction procedure, and Section 6.6 for the event reconstruction procedure.

The trigger used for the boosted selection in this analysis is the same as in the resolved selection, which uses single-lepton triggers. The previous analysis also used the large- $R$  jet triggers in the boosted selection, but this has been changed in this analysis for simplicity. In this analysis, an electron channel event is accepted by the trigger selection if either the isolated single-electron trigger with threshold at  $p_T > 24$  GeV or the non-isolated single-electron trigger with threshold at  $p_T > 60$  GeV are satisfied. In the muon channel, either the isolated single-muon trigger with threshold at  $p_T > 24$  GeV or the non-isolated single-muon trigger with threshold at  $p_T > 36$  GeV must be satisfied. The isolated lepton triggers show an efficiency loss at high transverse momentum, which is corrected for, by requiring that the higher  $p_T$  threshold non-isolated triggers are fulfilled alternatively <sup>1</sup>. The selected lepton is also required to match the trigger lepton, as previously, and scale factors are calculated based on the efficiency of these triggers in data and simulation to correct for discrepancies using the same procedure as described previously.

In the analysis performed with 2011 data, the jet closest to the lepton was taken as the  $b$ -jet from the leptonic top decay, while in this analysis the highest  $p_T$  jet which has a  $\Delta R(\text{jet}, \text{lepton}) < 1.5$  is used for the  $m_{t\bar{t}}$  reconstruction [105, 115]. It can be shown [115] that this choice improves the  $m_{t\bar{t}}$  resolution comparing the reconstructed result with the particle-level simulation value.

Another update is the usage of jet trimming [116] for the large- $R$  jets, which removes low energy clusters, reducing the effect of multiple interactions in the bunch crossing and initial state radiation. The jet trimming procedure reconstructs sub-jets inside the large- $R$  jet with a  $R_{\text{sub}} = 0.3$  parameter for the jet algorithm. If a sub-jet has a transverse momentum  $p_T < f_{\text{cut}}\Lambda_{\text{hard}}$ , then the sub-jet contribution to the large- $R$  jet four-momentum is discarded. In this analysis,  $f_{\text{cut}} = 0.05$  and  $\Lambda_{\text{hard}}$  is set to the transverse momentum of the original large- $R$  jet [115].

## 7.2 Multi-jet background modelling

The data-driven QCD multi-jets estimate is described in Section 5.5, but it was performed by the author for the  $\sqrt{s} = 8$  TeV search detailed in this chapter, therefore more information about the data-driven method and the calculation

---

<sup>1</sup>In the ATLAS jargon, it is required that either `EF_e24vhi_medium1` or `EF_e60_medium1` pass in the electron channel. In the muon channel, either `EF_mu24i_tight` or `EF_mu36_tight` must pass.

is detailed in this section. As was mentioned in Section 5.5, the QCD multi-jets contribution is estimated by defining a “loose” and a “tight” lepton definition and selecting real data considering the leptons in these two definitions (with the “loose” definition including the “tight” as a subset). Events that pass the selection performed with leptons in both the “tight” and “loose” configurations would be weighted by:

$$w_{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{eff}} - \epsilon_{\text{fake}}} \times (\epsilon_{\text{eff}} - 1), \quad (7.1)$$

while, events that only pass the selection with leptons in the “loose” configuration are weighted by:

$$w_{\text{loose}} = \frac{1}{\epsilon_{\text{eff}} - \epsilon_{\text{fake}}} \times (\epsilon_{\text{fake}} \epsilon_{\text{eff}}). \quad (7.2)$$

The important elements in the previous equations are the  $\epsilon_{\text{eff}}$  and  $\epsilon_{\text{fake}}$  terms, which should be estimated. The former gives the probability of a real lepton to pass the “tight” selection, given that it satisfies the “loose” selection. The latter, describes the probability of a jet to fake a “tight” lepton, by calculating the probability of an object to pass the “tight” lepton criteria, given that it satisfies the “loose” lepton criteria in a QCD multi-jets-enriched region.

The  $\epsilon_{\text{eff}}$  term should be calculated in a region enriched with true leptons. A typical calculation of this term involves defining a selection in real data for  $Z$ -boson lepton decays, for example, which would be enriched in electron or muon pairs in the  $Z$ -boson window mass region. The  $\epsilon_{\text{eff}}$  term could then be calculated using the Tag And Probe method, by selecting a lepton that satisfies the “loose” criteria in the designed selection and verifying how often another lepton that satisfies the “tight” criteria can be found that is still in the  $Z$ -boson mass window used. In this analysis, the lepton scale factors (derived themselves using the Tag And Probe method, as was described in [79] and [80]) show that there is a good agreement between the efficiencies for the “loose” and “tight” selections in simulation after using the lepton reconstruction scale factors and the data selection efficiencies within uncertainties, therefore the simulation can be used to estimate the  $\epsilon_{\text{eff}}$  factor. Events are selected in  $t\bar{t}$  simulation according to the standard selection used in the analysis with the extra requirement that there was in fact, at simulation level, a semileptonic  $t\bar{t}$  decay, which eliminates falsely identified events due to the detector effects. With these demands, the number of events satisfying the selection with the

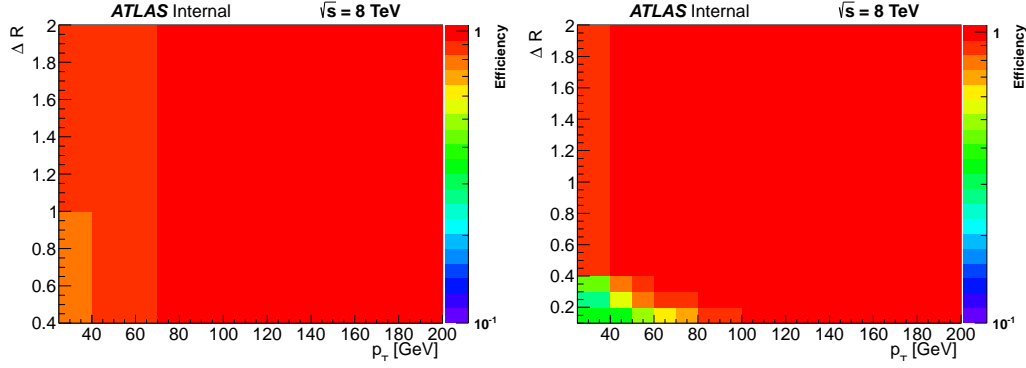


Figure 7.1:  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$  and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the electron (left) and muon (right) channels, for the resolved selection.

“tight” configuration divided by those satisfying the selection in the “loose” configuration is calculated.

The  $\epsilon_{\text{eff}}$  factor depends on kinematical variables of the event and a better estimate for the shape of the QCD multi-jet background can be calculated if this rate’s dependence on the relevant variables is taken into account. Ideally, a large set of the kinematical variables of the event could be used, but this would increase the statistical uncertainty of this variable in each bin, leading to a large uncertainty in the QCD multi-jet estimate. In this analysis, the  $\epsilon_{\text{eff}}$  variable shows a large dependence on the lepton transverse momentum and the smallest  $\Delta R$  between the lepton and a small- $R$  jet. These quantities were used jointly to parametrise  $\epsilon_{\text{eff}}$ . Figure 7.1 shows the parametrization used for the resolved selection in the electron and muon channels. In the boosted channel, due to the smaller number of events, the two dimensional parametrization is only used in the muon channel for the  $\Delta R \leq 0.4$  situation and it is shown in Figure 7.2. For the electron channel, in which the  $\Delta R$  variable is always greater than 0.4, and for the muon channel if  $\Delta R > 0.4$ , the parametrization as a function of the lepton transverse momentum only is used and it is shown in Figure 7.3.

The  $\epsilon_{\text{fake}}$  variable should be calculated in a QCD multi-jets enriched region. A control region is designed to enhance the QCD multi-jet contribution in data and reduce the other background contributions. In this control region, the data can be subtracted from the backgrounds. This subtraction procedure can be done for events that satisfy the “tight” selection or the “loose” selection, so that their ratio can be calculated to determine  $\epsilon_{\text{fake}}$ .

The control region is defined by changing the standard selection in this anal-

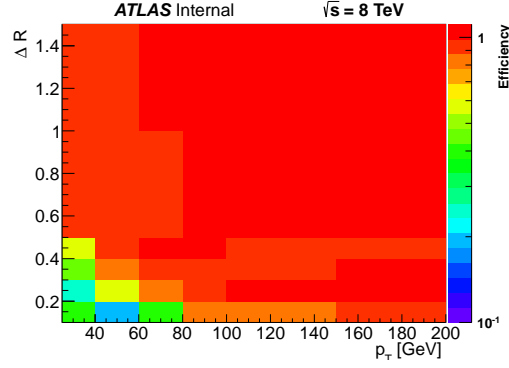


Figure 7.2:  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$  and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the muon channel, for the boosted selection. In the muon channel, to reduce the statistical uncertainty, this parametrisation is only used for muons with min ( $\Delta R(\text{lepton}, \text{jet})$ )  $\leq 0.4$  and a parametrisation solely described by the muon  $p_T$  is used otherwise.

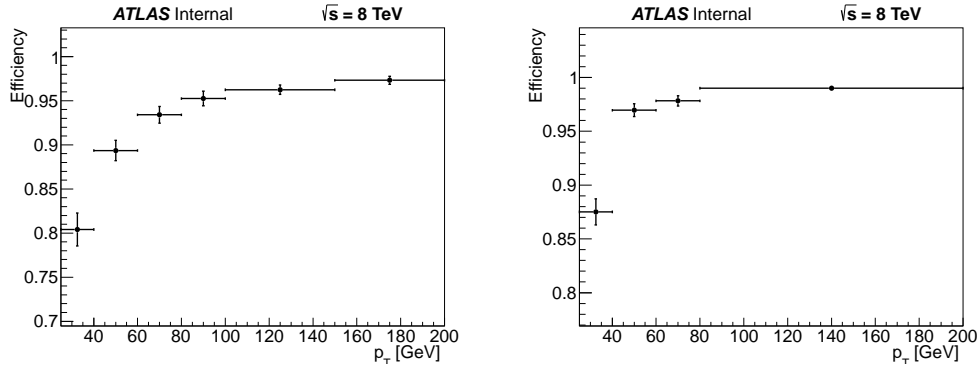


Figure 7.3:  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$  for the electron (left) and muon (right) channels, in the boosted selection, which is used if the min( $\Delta R(\text{lepton}, \text{jet})$ )  $> 0.4$ . In the muon channel, the previous criteria might not be satisfied and a parametrisation in function of both these variables is used in such a case.

ysis in only a few requirements. For the resolved selection, the missing transverse energy requirement and the transverse mass of the lepton and missing transverse energy requirement are both inverted. While in the standard selection, it is required  $E_T^{\text{miss}} > 30$  GeV and  $m_T > 30$  GeV for the electron channel, and, for the muon channel,  $E_T^{\text{miss}} > 20$  GeV and  $m_T > 60$  GeV  $- E_T^{\text{miss}}$ ; in the (background subtracted) control region, the requirement is  $E_T^{\text{miss}} \leq 30$  GeV and  $m_T \leq 30$  GeV for the electron channel, and, for the muon channel,  $E_T^{\text{miss}} \leq 20$  GeV and  $m_T \leq 60$  GeV  $- E_T^{\text{miss}}$ . In the boosted selection, the transverse mass and transverse energy requirements are changed so that, in both electron and muon channels it is demanded that  $E_T^{\text{miss}} \leq 60$  GeV; for the electron channel it is demanded that  $m_T \leq 60$  GeV and for the muon channel,  $m_T \leq 60$  GeV  $- E_T^{\text{miss}}$ . The requirements in the boosted selection are loosened to decrease the statistical uncertainty. The large- $R$  jet requirements are changed so that no large- $R$  jets are found with the same mass,  $\Delta R$  and  $\Delta\Phi$  selection requirements as in the signal region, but the large- $R$  jet transverse momentum requirement is relaxed so that it is  $p_T > 150$  GeV, and no requirement is made on the first  $k_t$  splitting scale,  $\sqrt{d_{12}}$ . This requirement in the control region inverts and tightens the ones of the signal region, so that the selections are orthogonal to each other, while it is still permitted for lower mass or transverse momentum jets to be found. The selected large- $R$  jet to represent the hadronic top quark in the control region is redefined to be the highest transverse momentum large- $R$  jet available <sup>2</sup>.

The Control Region definition includes a further requirement: the modulus of the transverse impact parameter significance ( $S(d_0)$ ) for the lepton track must be greater than 2.5 in the electron channel and greater than 4, in the muon channel. This requirement is not essential to the analysis, but it increases the fraction of heavy flavour in the Control Region sample. A dependency of the fake rate on the heavy flavour content of the QCD multi-jets sample would then be clearer. This effect has been seen previously in an analysis that does not require many selection cuts in Chapter 4, particularly in Figure 4.2.

The effect of the  $|S(d_0)|$  requirement, can be seen to enhance the heavy flavour in this analysis, calculating the fraction of  $b$ -tagged events in the analysis for different requirements for  $|S(d_0)|$  cuts, as shown in Figure 7.4. In this figure, the Control Region requirements are relaxed, by not applying the  $S(d_0)$  selection and not demanding the  $b$ -tagging selection, while all other require-

---

<sup>2</sup>This choice is not important for the QCD parametrisation, since it is not used for that end. The hadronic top in the boosted regime control region is only used to calculate the  $m_{t\bar{t}}$  variable for cross checks, as it is shown in what follows.



ments are kept. This includes all events that pass the loose criteria. It can be seen that for higher values of the  $|S(d_0)| > \text{cut}$  requirement, the fraction of  $b$ -tagged events increases. Figure 7.5 shows the fraction of events in each  $|S(d_0)|$  bin that have a certain number of jets that passed the  $b$ -tagging criteria. It is normalised to have a sum of number of  $b$ -tagged jets entries equal to one for a fixed  $|S(d_0)|$ .

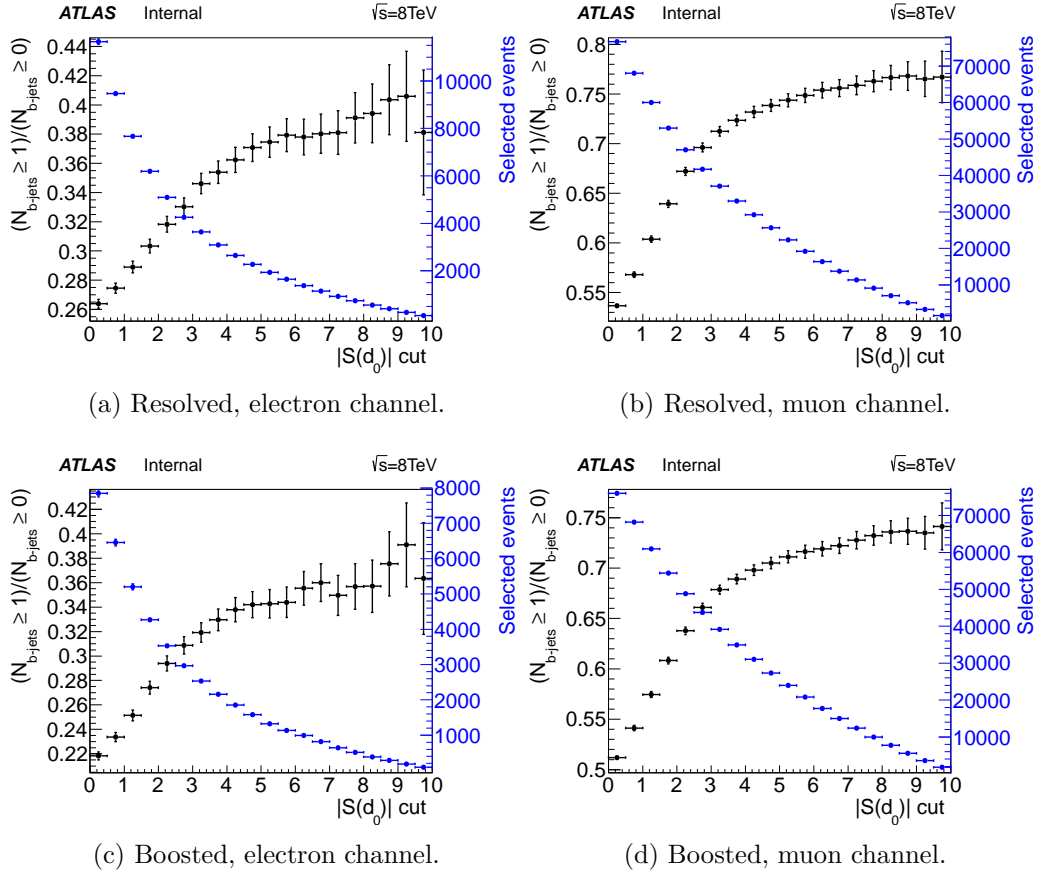


Figure 7.4: The number of  $b$ -tagged events over all events in the Control Region. For these plots, no  $S(d_0)$  and  $b$ -tagging cut were required for all events. The loose criteria is required.

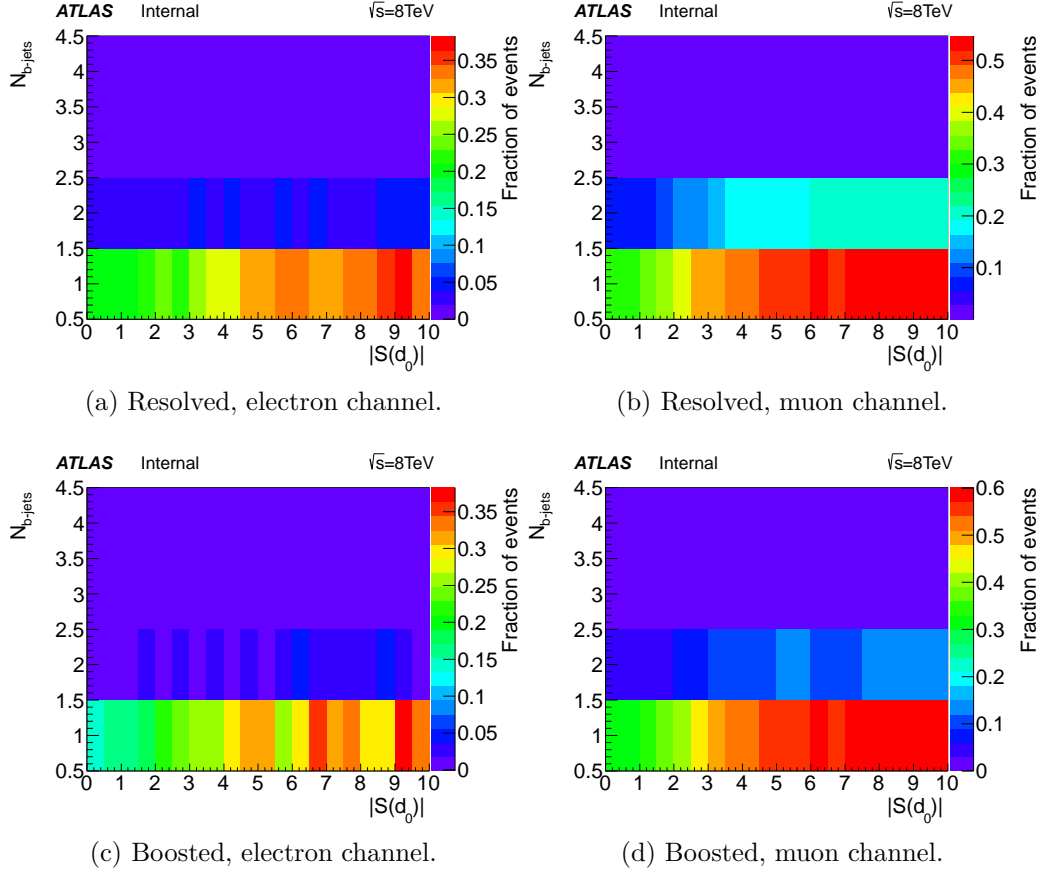


Figure 7.5: The fraction of  $b$ -tagged jets versus the  $|S(d_0)|$  of the event in the Control Region. For these plots, no  $S(d_0)$  and  $b$ -tagging cut were required for all events. The loose criteria is required.

The choice of variables used in the parametrisation of  $\epsilon_{\text{fake}}$  is such that the fake rate has a large dependence on them. In both resolved and boosted selections, there are three relevant variables that are used: the smallest  $\Delta R$  between the lepton and a jet, the lepton transverse momentum and the transverse momentum of the jet closest to the lepton. For the electron channel, only the latter two variables are used. The muon channel is parametrised similarly, but events are separated in two categories, depending on whether the  $\Delta R$  variable is greater than or smaller than 0.4. The  $\epsilon_{\text{fake}}$  parametrisation is shown in Figures 7.6, 7.7 and 7.8.

The systematic uncertainties for this parametrisation were calculated, considering the electron and muon scale factor systematic uncertainty <sup>3</sup>, the  $b$ -

<sup>3</sup>The electron and muon scale factors are the ratio between the efficiency of their selection in data over the efficiency of their selection in simulation. The efficiency in data was calculated using the Tag And Probe method (see Chapter 3 and Section 5.6). The scale factor was used for the nominal value of efficiencies and fake rates.

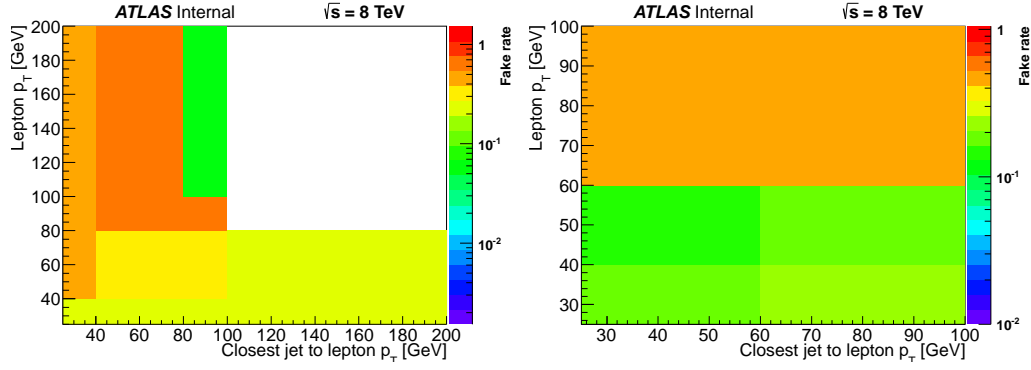


Figure 7.6:  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the electron (left) and muon (right) channels, in the resolved selection, only for  $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ .

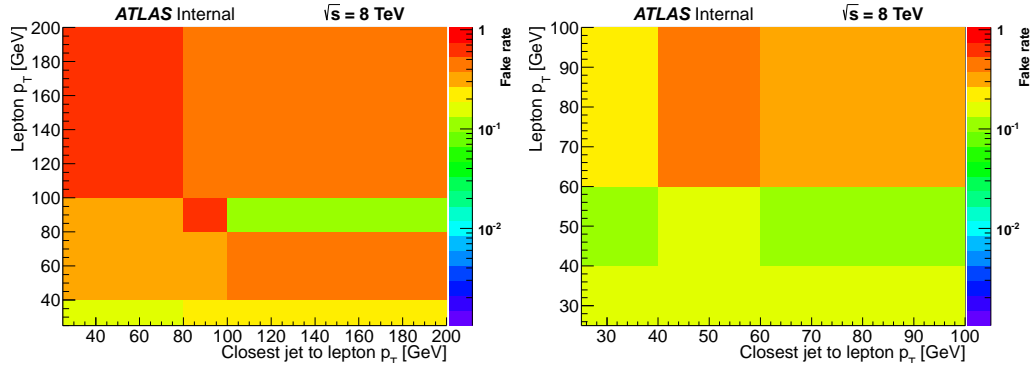


Figure 7.7:  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, only for  $\min(\Delta R(\text{lepton}, \text{jet})) > 0.4$ .

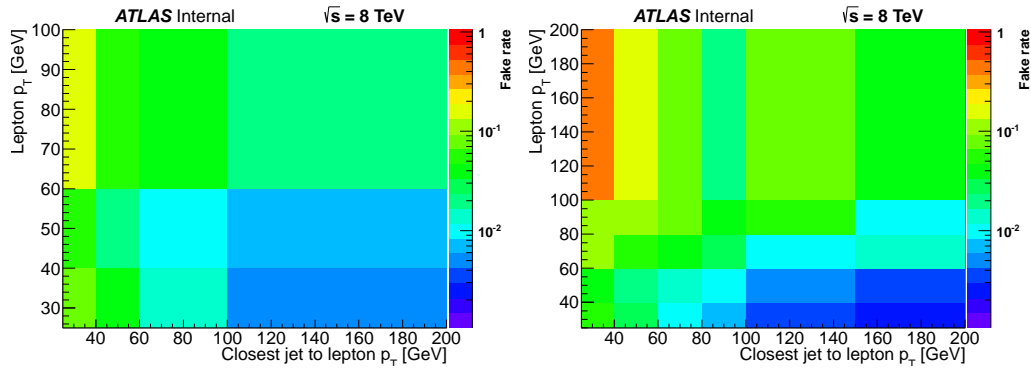


Figure 7.8:  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the muon channel, in the resolved selection (left) and boosted selection (right), only for  $\min(\Delta R(\text{lepton}, \text{jet})) \leq 0.4$ .

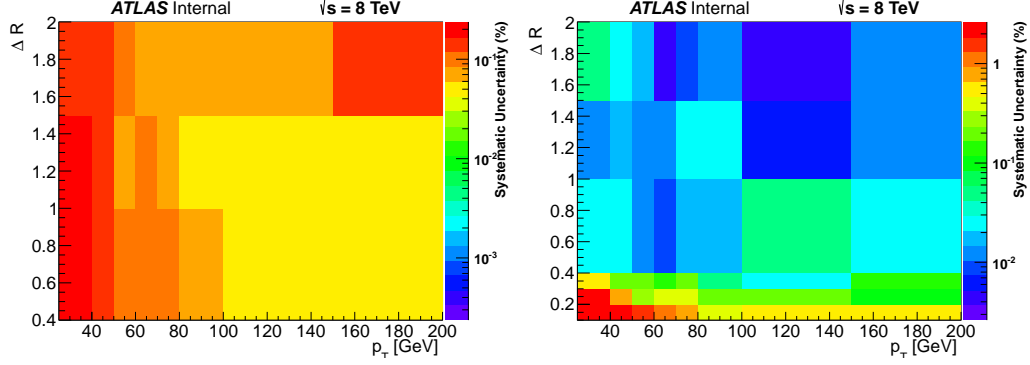


Figure 7.9: Systematic uncertainty in  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$  and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the electron (left) and muon (right) channels, for the resolved selection.

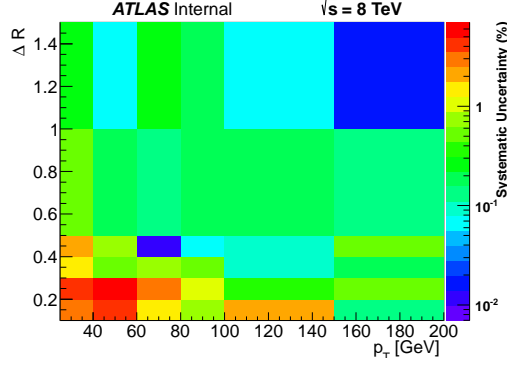


Figure 7.10: Systematic uncertainty in  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$  and the min ( $\Delta R(\text{lepton}, \text{jet})$ ) in the muon channel, for the boosted selection. In the muon channel, to reduce the statistical uncertainty, this parametrisation is only used for muons with min ( $\Delta R(\text{lepton}, \text{jet})$ )  $\leq 0.4$  and a parametrisation solely described by the muon  $p_T$  is used otherwise.

tagging systematic uncertainty and the jet vertex fraction systematic uncertainty. The percentage variation in each bin is shown in Figures 7.9, 7.10, 7.11 for the efficiency in the signal region and in Figures 7.12, 7.13 and 7.14 for the fake rate. It can be seen that there are a few bins in which the uncertainties are very big (note particularly, bin at (90 GeV, 150 GeV) in Figure 7.12 and the bin (150 GeV, 90 GeV) in Figure 7.13, in the electron channel), but for those bins the statistical uncertainty is already very large ( $\sim 1500\%$  for the former,  $\sim 300\%$  for the latter) and any small variation is enough to cause such a large relative shift. The uncertainties in the mentioned bins do not affect the QCD prediction significantly, since the amount of real data (in the loose and tight selections) in these bins, for the signal region is small (0.98% in the former and 3.91% in the latter).

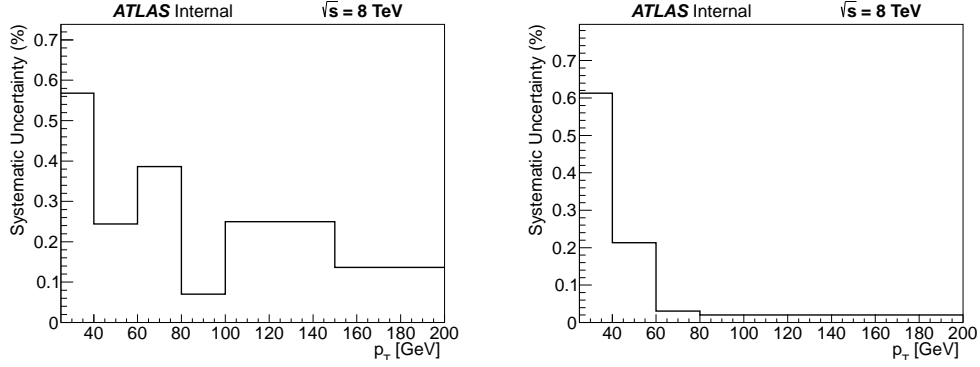


Figure 7.11: Systematic uncertainty in  $\epsilon_{\text{eff}}$  parametrised as a function of the lepton  $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, which is used if the min ( $\Delta R(\text{lepton}, \text{jet})$ )  $> 0.4$ . In the muon channel, the previous criteria might not be satisfied and a parametrisation as a function of both these variables is used in such a case.

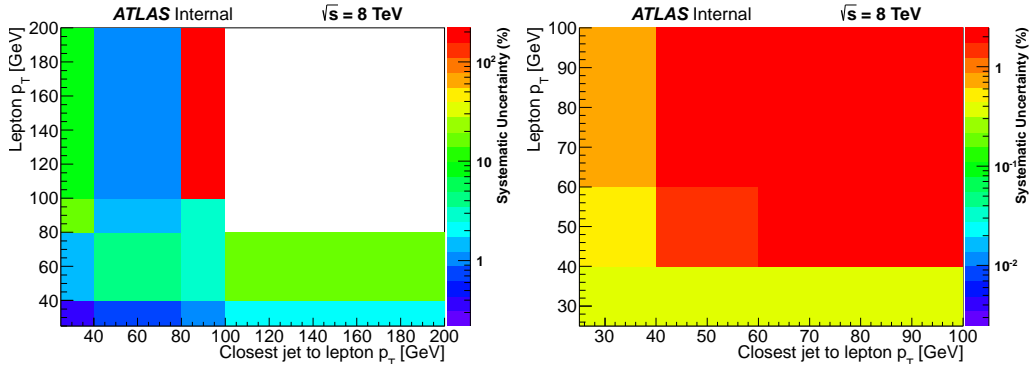


Figure 7.12: Systematic uncertainty in  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the electron (left) and muon (right) channels, in the resolved selection, only for min ( $\Delta R(\text{lepton}, \text{jet})$ )  $> 0.4$ .

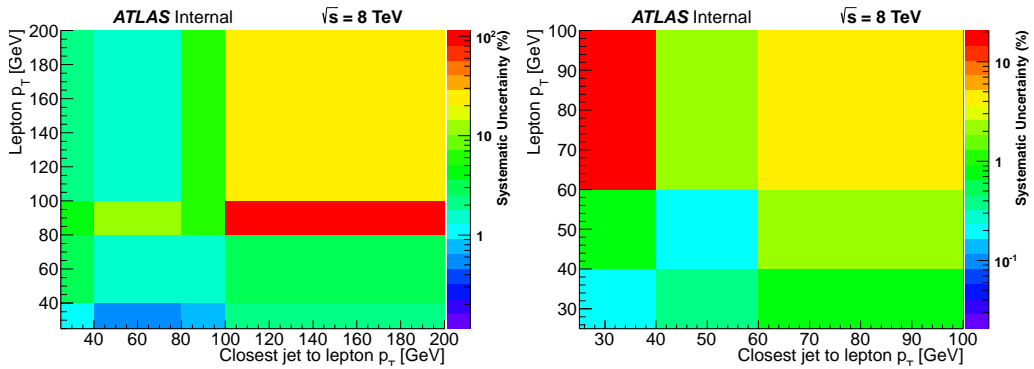


Figure 7.13: Systematic uncertainty in  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the electron (left) and muon (right) channels, in the boosted selection, only for min ( $\Delta R(\text{lepton}, \text{jet})$ )  $> 0.4$ .

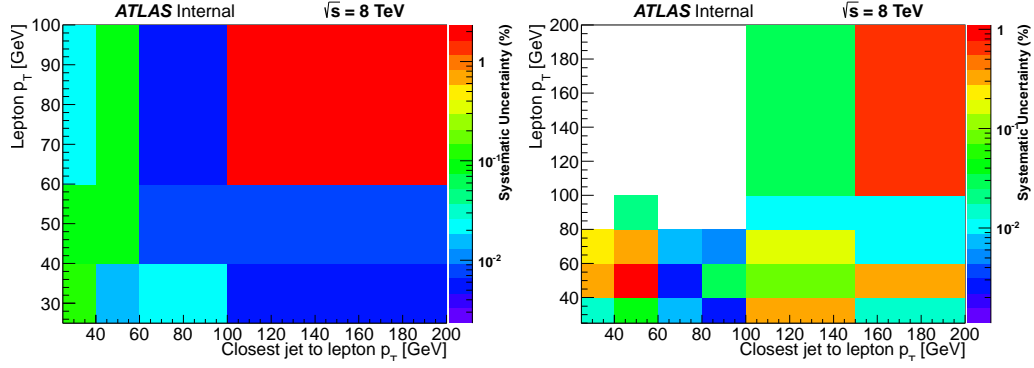


Figure 7.14: Systematic uncertainty in  $\epsilon_{\text{fake}}$  parametrised as a function of the lepton  $p_T$  and the closest jet to lepton  $p_T$ , for the muon channel, in the resolved selection (left) and boosted selection (right), only for  $\min(\Delta R(\text{lepton}, \text{jet})) \leq 0.4$ .

With these parametrisations and the weighting mechanism from Equations 7.1 and 7.2 (with details in Section 5.5) a consistency check can be done, by calculating the  $m_{t\bar{t}}$  variable in the control region used to calculate the  $\epsilon_{\text{fake}}$  variable, which is enriched in the QCD multi-jets background. The  $m_{t\bar{t}}$  variable calculated in data, in the backgrounds and the QCD multi-jets estimate result can be seen in Figures 7.15 and 7.16. There is good agreement between the data and the estimated QCD multi-jets backgrounds, which shows that the method works reasonably well in estimating these variables. The systematic uncertainty included for the multi-jets sample in these plots comes from the propagation of the estimated systematic uncertainties of the  $\epsilon_{\text{eff}}$  and  $\epsilon_{\text{fake}}$  terms mentioned previously added in quadrature with their statistical uncertainty. The systematic uncertainties for other backgrounds were estimated in the same way as the rest of the analysis. The total uncertainty due to the systematic variation in the  $\epsilon_{\text{eff}}$  term, in the resolved channel, is 0.04% in both electron and muon channels, and it is 0.22% in the boosted electron channel, and 0.46% in the boosted muon channel. The variation in the  $\epsilon_{\text{fake}}$  term, amounts to a systematic uncertainty of 15.97% in the resolved electron channel, 7.26% in the resolved muon channel, 15.53% in the boosted electron channel and 15.34% in the boosted muon channel. Nonetheless, the systematic uncertainty associated to this background in the analysis results mentioned in the next section relate to a comparison of this method and other methods used to estimate the QCD multi-jets background (with other parametrisation, control region and loose lepton definitions), which amounts to a conservative 50% normalisation uncertainty.

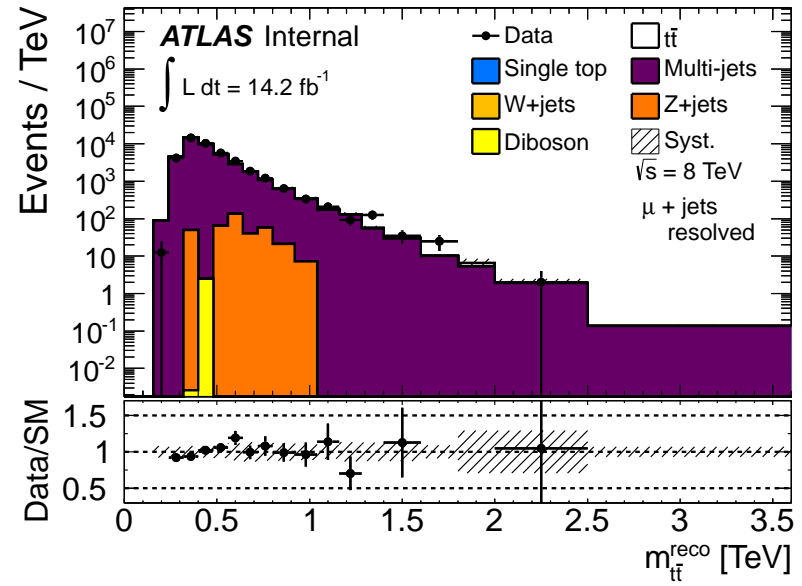
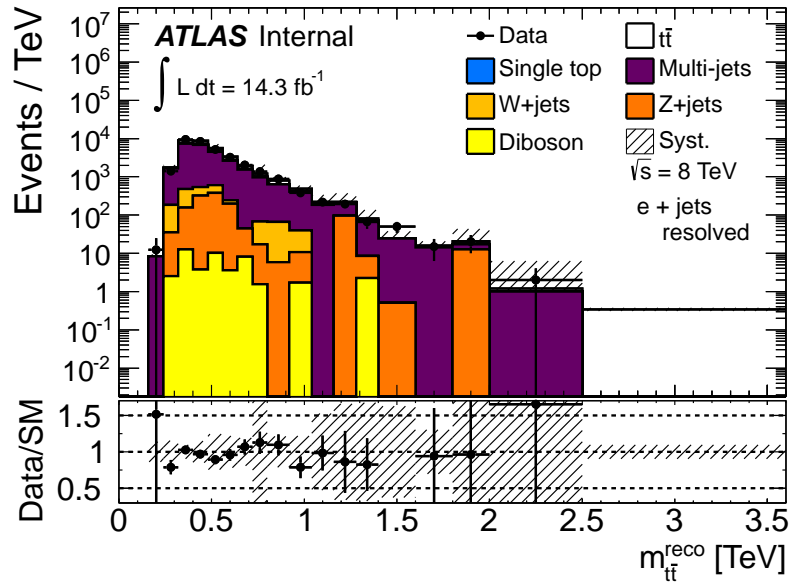


Figure 7.15:  $m_{t\bar{t}}$  variable calculated in the resolved scenario, in the QCD multi-jets enriched control region, for the electron (left) and muon (right) channels.

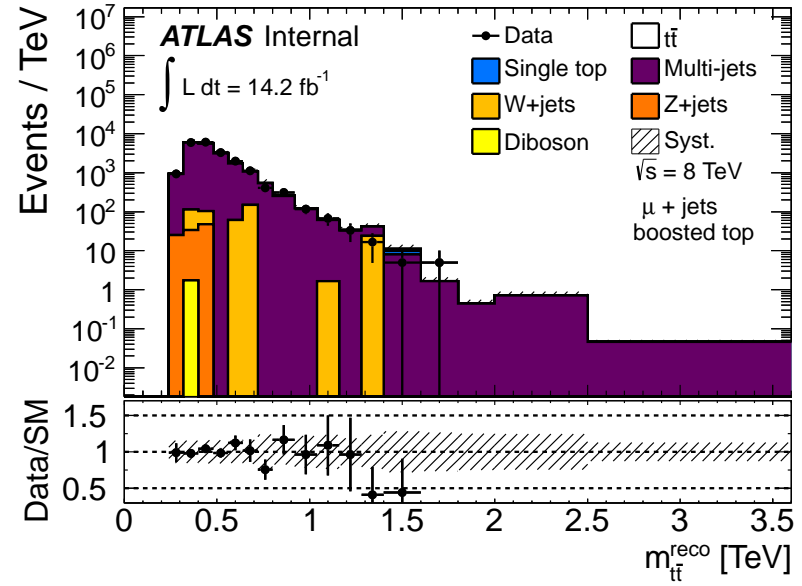
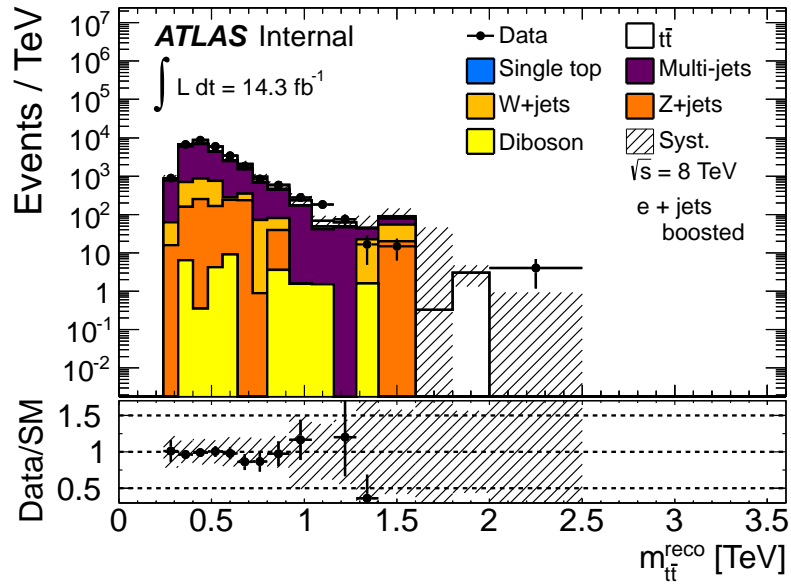


Figure 7.16:  $m_{t\bar{t}}$  variable calculated in the boosted scenario, in the QCD multi-jets enriched control region, for the electron (left) and muon (right) channels.



## 7.3 Event reconstruction and results

The event reconstruction follows a similar procedure to the one described in the previous chapter, with a small difference in the selection of the jet from the leptonically decaying top quark in the boosted selection, as mentioned previously. The total event count in each channel is shown in Tables 7.1, 7.2, 7.3 and 7.4. The systematic variations in the data and expectations are shown in Tables 7.5, 7.6, 7.7 and 7.8.

Table 7.1: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 8$  TeV in the resolved electron channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds.

Source	Yield
$t\bar{t}$	$93390 \pm 194 \pm 18806$
single top	$6764 \pm 73 \pm 1462$
QCD multi-jets	$3681 \pm 86 \pm 1840$
W+jets	$15633 \pm 382 \pm 6051$
Z+jets	$1790 \pm 72 \pm 1352$
Diboson	$225 \pm 7 \pm 128$
Expectation	$121485 \pm 449 \pm 30029$
Data	$119520 \pm 345$

Table 7.2: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 8$  TeV in the resolved muon channel with statistical uncertainties for the data and background samples, followed by the total systematic uncertainty for the backgrounds.

Source	Yield
$t\bar{t}$	$117700 \pm 218 \pm 23417$
single top	$8411 \pm 82 \pm 1929$
QCD multi-jets	$10475 \pm 53 \pm 5237$
W+jets	$23047 \pm 485 \pm 8582$
Z+jets	$1793 \pm 75 \pm 1278$
Diboson	$318 \pm 8 \pm 152$
Expectation	$161747 \pm 546 \pm 41157$
Data	$160997 \pm 401$

The data to simulation comparison for the analysis is included in the plots in Figures 7.17, 7.18, 7.19, 7.20 and 7.21. Figure 7.17 shows the transverse momentum of the leading jet in the resolved scenario. Figure 7.18 shows the transverse momentum of the hadronically decaying top quark candidate in the boosted selection, which is chosen as the highest  $p_T$  large- $R$  jet that satisfies

Table 7.3: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 8$  TeV in the boosted electron channel with statistical uncertainties for data and background samples, followed by the total systematic uncertainty for the background samples.

Source	Yield
$t\bar{t}$	$2083 \pm 28 \pm 588$
single top	$70 \pm 6 \pm 34$
QCD multi-jets	$39 \pm 11 \pm 19$
W+jets	$167 \pm 31 \pm 123$
Z+jets	$18 \pm 7 \pm 53$
Diboson	$2 \pm 0 \pm 2$
Expectation	$2381 \pm 45 \pm 1026$
Data	$2178 \pm 46$

Table 7.4: Total contribution of each of the background samples in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 8$  TeV in the boosted muon channel with the statistical uncertainties for data and background samples, followed by the total systematic uncertainty for the background samples.

Source	Yield
$t\bar{t}$	$2785 \pm 33 \pm 769$
single top	$105 \pm 7 \pm 42$
QCD multi-jets	$32 \pm 2 \pm 16$
W+jets	$308 \pm 47 \pm 190$
Z+jets	$33 \pm 10 \pm 53$
Diboson	$1 \pm 0 \pm 3$
Expectation	$3266 \pm 59 \pm 1329$
Data	$2945 \pm 54$

the selection cuts. Figure 7.19 shows the mass of the leptonically decaying top in the boosted scenario, which is reconstructed adding the four-momenta of the neutrino, the lepton and the jet selected as the  $b$ -jet from the top quark leptonic decay. The  $b$ -jet from the top quark leptonic decay, as described previously, is chosen as the highest transverse momentum small- $R$  jet which has a  $\Delta R(\text{jet}, \text{lepton}) < 1.5$ . Figure 7.20 shows the mass of the hadronically decaying top quark candidate. Figure 7.21 shows the first  $k_t$  splitting scale,  $\sqrt{d_{12}}$ , for the selected large- $R$  jet taken as the hadronically decaying top quark. The final spectra are given in Figures 7.22 for the resolved selection and 7.23, for the boosted selection. These spectra with all systematics are used to set the limits for the analysis. Figure 7.24 shows the sum of all four channels in a single histogram, with one of the invariant mass configurations for each of the benchmark models overlayed for illustration (with their production cross

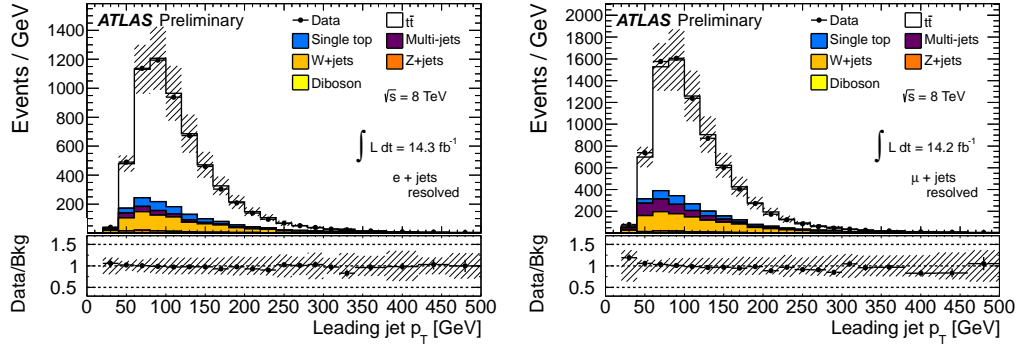
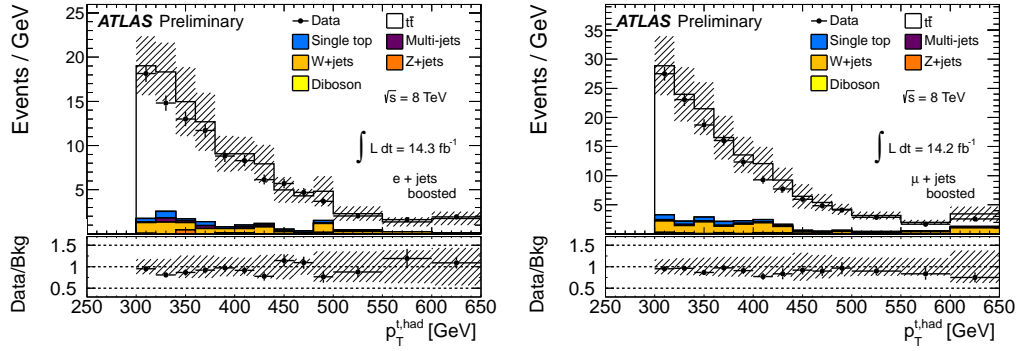


Figure 7.17: Transverse momentum of the leading jet in the resolved scenario.

Figure 7.18: Transverse momentum of the large- $R$  jet chosen as the hadronically decaying top quark candidate in the boosted selection.

section multiplied by five).

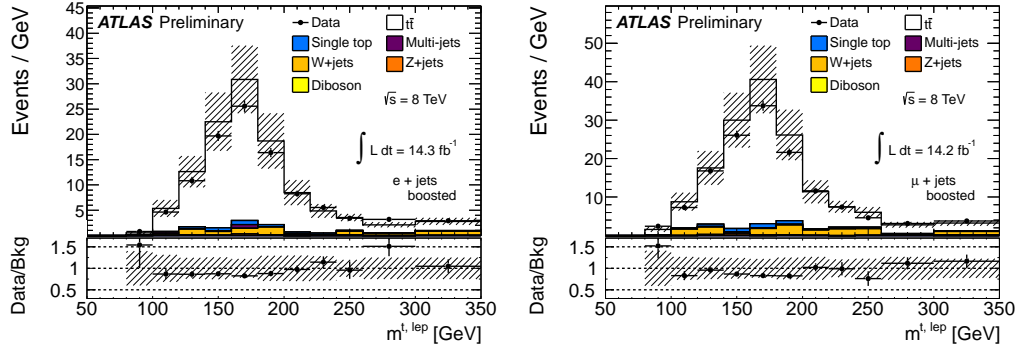


Figure 7.19: Invariant mass of the leptonically decaying top quark candidate in the boosted selection.

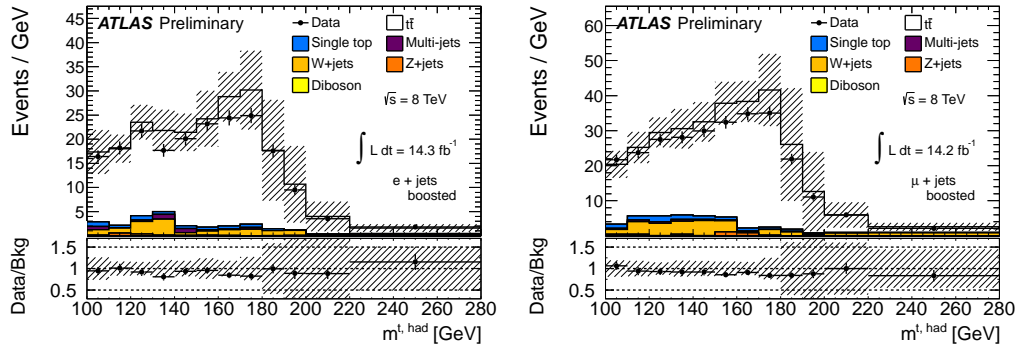


Figure 7.20: Mass of the large- $R$  jet chosen as the hadronically decaying top quark candidate in the boosted selection.

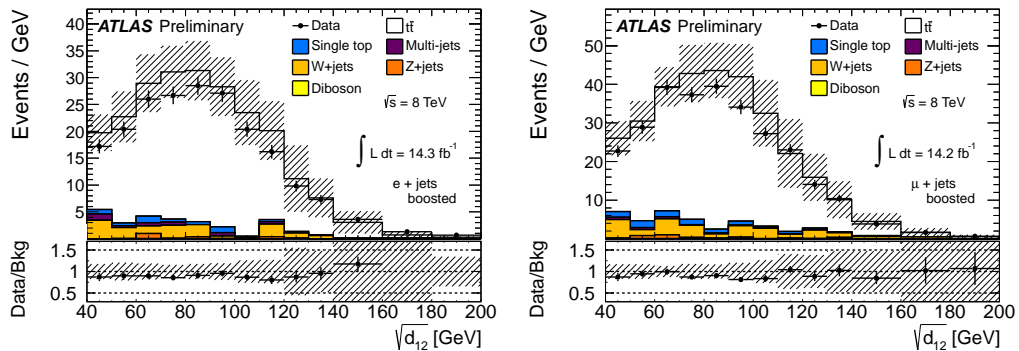
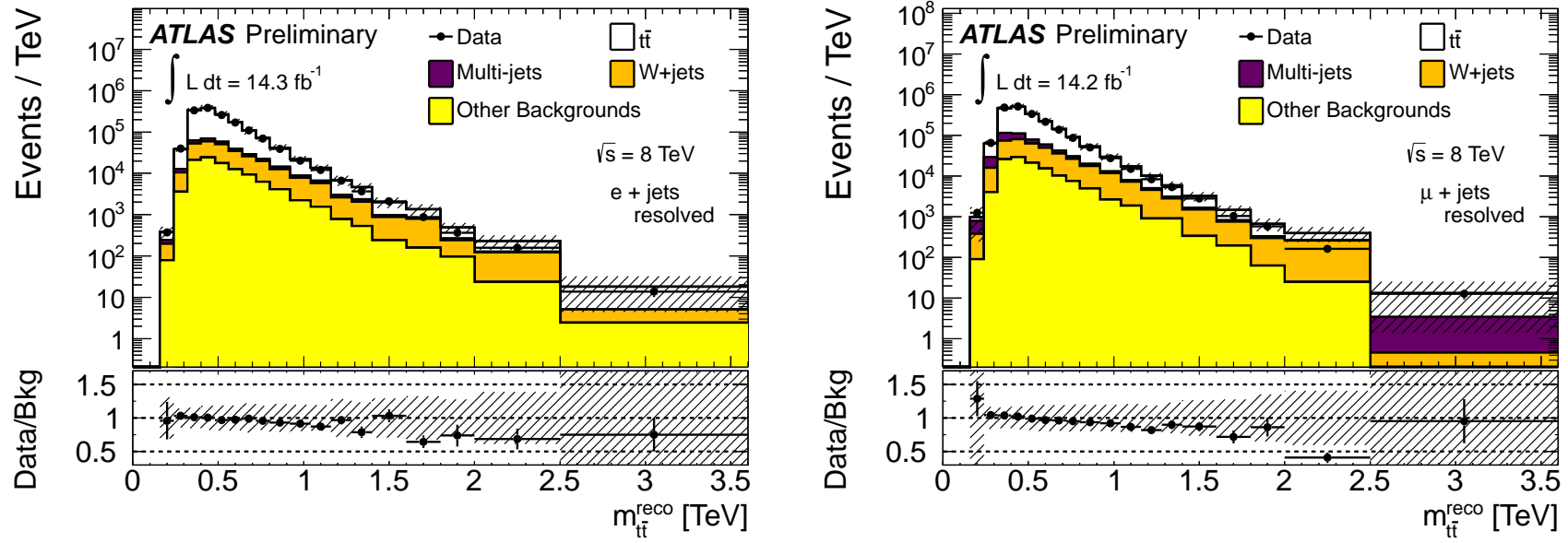
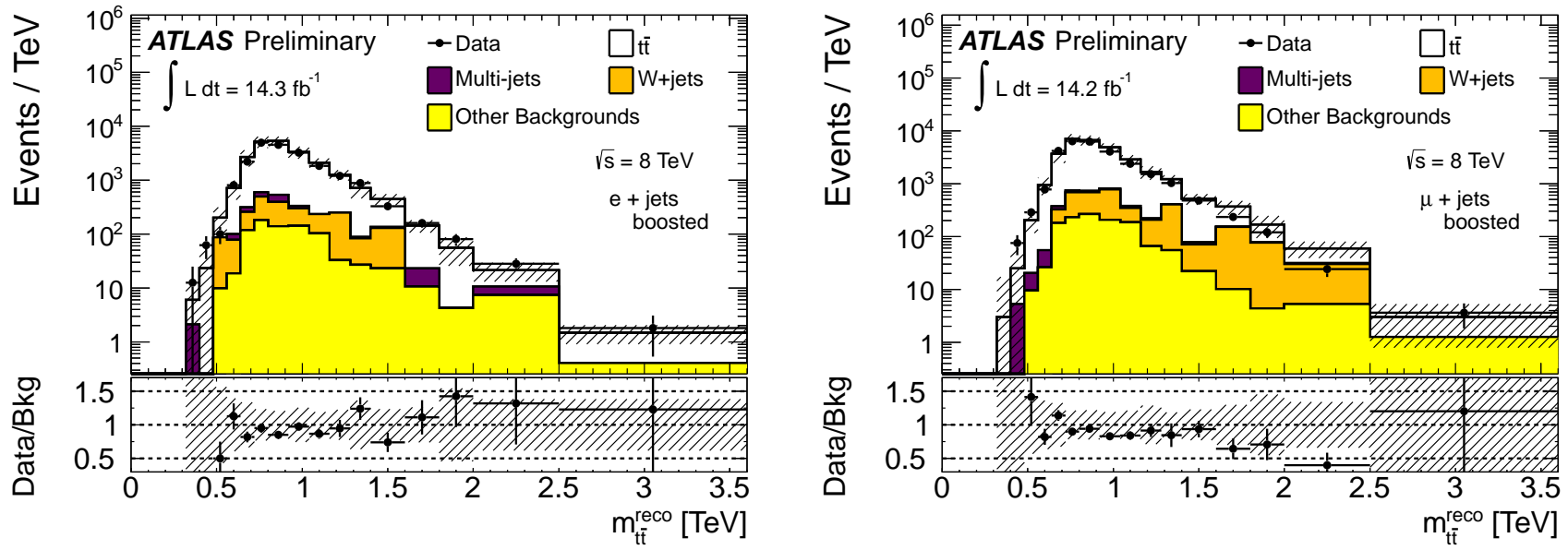


Figure 7.21: First splitting scale,  $\sqrt{d_{12}}$  for the large- $R$  jet chosen as the hadronically decaying top quark candidate in the boosted selection.

Figure 7.22: Reconstructed invariant mass of the  $t\bar{t}$  system in the resolved scenario.

Figure 7.23: Reconstructed invariant mass of the  $t\bar{t}$  system in the boosted scenario.

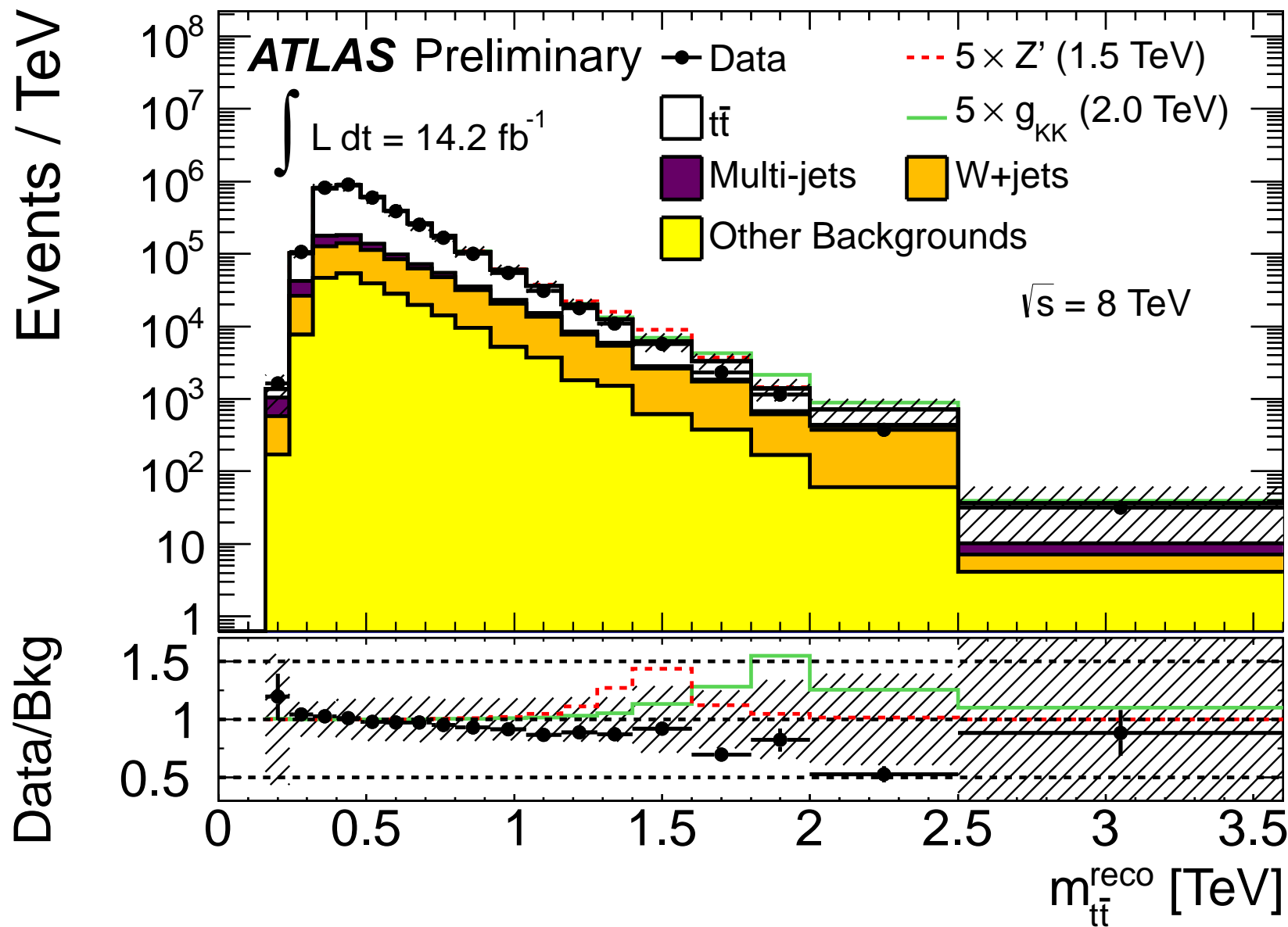


Figure 7.24: Reconstructed invariant mass of the  $t\bar{t}$  system for the resolved, boosted, electron and muon channels summed in a single histogram. One mass point for each benchmark model in the analysis is overlaid with the background, having their production cross section multiplied by five.

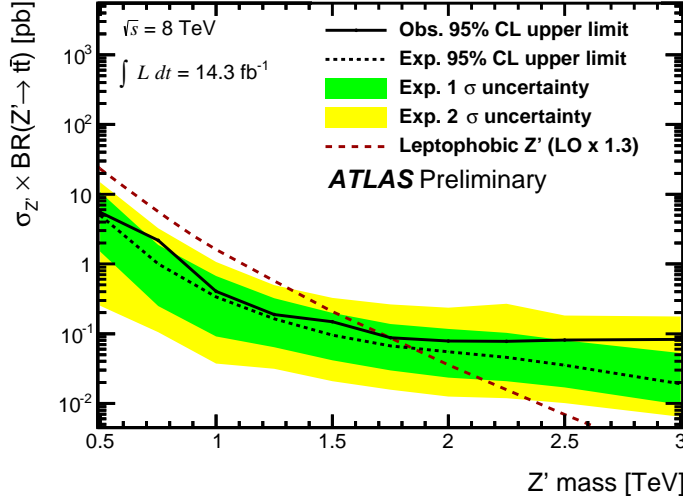


Figure 7.25: Observed and expected upper cross section times branching ratio limit for a narrow  $Z'$  resonance. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a  $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [115].

## 7.4 Limit setting and summary

The signal and backgrounds were estimated in four channels, for the boosted and resolved, electron and muon channels, in the  $\sqrt{s} = 8$  TeV  $t\bar{t}$  resonances search analysis in a very similar way as it was done in the  $\sqrt{s} = 7$  TeV search. All major systematic uncertainties were considered and estimated.

As in the  $\sqrt{s} = 7$  TeV result, the spectra information calculated with all its systematic and statistical uncertainties were used to test the hypothesis that the benchmark models are valid. The BumpHunter tool [113] was used, but no significant deviation from the Standard Model prediction was found. The Bayesian limit setting procedure described in [114] was implemented to set limits on the parameters of the models and the results are summarised in Figures 7.25 and 7.26. The  $Z'$  mass between 0.5 TeV and 1.8 TeV and the Kaluza-Klein gluon mass between 0.5 TeV and 2.0 TeV are excluded with 95% Confidence Level.



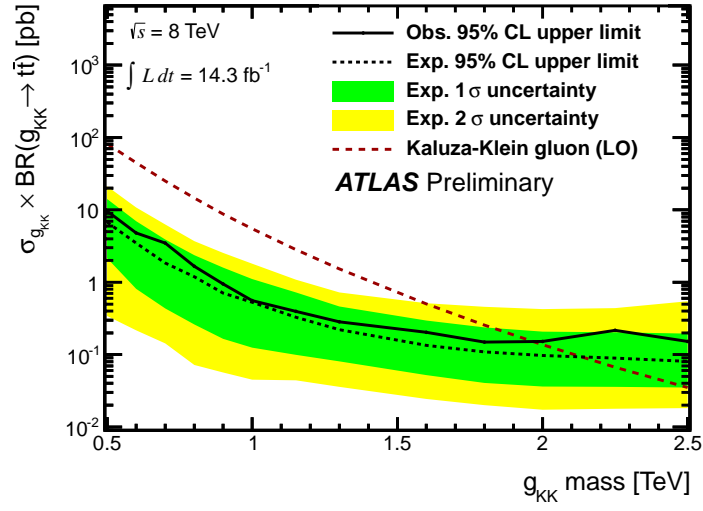


Figure 7.26: Observed and expected upper cross section times branching ratio limit for a Kaluza-Klein gluon. The resolved and boosted scenarios were combined. The red dotted line shows the theoretical cross section times branching ratio for the resonance with a  $k$ -factor that corrects its normalisation from the leading-order estimate to the next-to-leading order one. Extracted from [115].

Table 7.5: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the resolved electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	4.64
b-tagging c-tag rate	1.32
mistag rate	0.75
Jet efficiency	0.13
Jet energy resolution	2.13
Jet energy scale	12.62
Large- $R$ JES/JMS	0.34
JVF	2.08
W+jets shape	0.48
Z+jets norm.	0.74
W+jets bb+cc+c versus light	0.94
W+jets bb+25%	0.40
W+jets c+25%	1.23
W+jets norm.	2.06
ISR/FSR	0.13
Parton shower	4.58
Top mass	1.79
MC generator	1.44
$t\bar{t}$ electroweak Sudakov corr.	2.27
PDF syst.	2.99
$t\bar{t}$ cross section norm.	8.23
Diboson norm.	0.06
Single top norm.	0.43
Electron id./rec.	2.31
Electron energy scale	0.95
Electron energy resolution	0.26
Muon id./rec.	0.00
Muon momentum resolution (MS)	0.00
Muon momentum resolution (ID)	0.00
Muon momentum scale	0.02
$E_T^{\text{miss}}$ Soft Jet scale	0.41
$E_T^{\text{miss}}$ Soft Jet resolution	0.36
Luminosity	3.60
DD QCD norm.	1.52
All syst.	24.77
Stat. MC	0.37
Stat. Data	0.28

Table 7.6: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the resolved muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	4.48
b-tagging c-tag rate	1.45
mistag rate	0.75
Jet efficiency	0.14
Jet energy resolution	2.03
Jet energy scale	11.56
Large- $R$ JES/JMS	0.32
JVF	2.15
W+jets shape	0.56
Z+jets norm.	0.56
W+jets bb+cc+c versus light	0.89
W+jets bb+25%	0.15
W+jets c+25%	0.97
W+jets norm.	2.14
ISR/FSR	0.21
Parton shower	4.52
Top mass	1.74
MC generator	1.53
$t\bar{t}$ electroweak Sudakov corr.	2.15
PDF syst.	2.84
$t\bar{t}$ cross section norm.	7.79
Diboson norm.	0.07
Single top norm.	0.40
Electron id./rec.	0.00
Electron energy scale	0.03
Electron energy resolution	0.01
Muon id./rec.	2.56
Muon momentum resolution (MS)	0.07
Muon momentum resolution (ID)	0.05
Muon momentum scale	0.17
$E_T^{\text{miss}}$ Soft Jet scale	0.37
$E_T^{\text{miss}}$ Soft Jet resolution	0.38
Luminosity	3.60
DD QCD norm.	3.24
All syst.	25.65
Stat. MC	0.34
Stat. Data	0.25

Table 7.7: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis at  $\sqrt{s} = 8$  TeV, in the boosted electron channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	5.23
b-tagging c-tag rate	0.93
mistag rate	0.25
Jet efficiency	0.10
Jet energy resolution	1.83
Jet energy scale	8.62
Large- $R$ JES/JMS	19.71
JVF	2.09
W+jets shape	1.17
Z+jets norm.	0.42
W+jets bb+cc+c versus light	1.31
W+jets bb+25%	0.57
W+jets c+25%	0.90
W+jets norm.	0.95
ISR/FSR	0.84
Parton shower	3.68
Top mass	0.93
MC generator	0.71
$t\bar{t}$ electroweak Sudakov corr.	4.53
PDF syst.	5.61
$t\bar{t}$ cross section norm.	9.36
Diboson norm.	0.03
Single top norm.	0.23
Electron id./rec.	2.33
Electron energy scale	1.60
Electron energy resolution	0.99
Muon id./rec.	0.00
Muon momentum resolution (MS)	0.01
Muon momentum resolution (ID)	0.01
Muon momentum scale	0.14
$E_T^{\text{miss}}$ Soft Jet scale	1.30
$E_T^{\text{miss}}$ Soft Jet resolution	1.64
Luminosity	3.60
DD QCD norm.	0.89
All syst.	43.24
Stat. MC	1.92
Stat. Data	1.97

Table 7.8: Systematic uncertainties from all backgrounds in percentage variation of the  $t\bar{t}$  sample, in the  $t\bar{t}$  resonances analysis, in the boosted muon channel, using the maximum between the up and down variations. Total effect estimated in the yield of the background samples (no bin width weight applied).

Syst.	Percentage variation
b-tagging eff.	4.03
b-tagging c-tag rate	0.72
mistag rate	1.04
Jet efficiency	0.10
Jet energy resolution	2.33
Jet energy scale	8.16
Large- $R$ JES/JMS	18.55
JVF	2.20
W+jets shape	2.12
Z+jets norm.	0.56
W+jets bb+cc+c versus light	1.10
W+jets bb+25%	0.78
W+jets c+25%	0.67
W+jets Norm.	1.09
ISR/FSR	0.59
Parton shower	4.37
Top mass	1.60
MC generator	2.33
$t\bar{t}$ electroweak Sudakov corr.	4.45
PDF syst.	5.83
$t\bar{t}$ cross section norm.	9.13
Diboson norm.	0.02
Single top norm.	0.25
Electron id./rec.	0.00
Electron energy scale	0.08
Electron energy resolution	0.04
Muon id./rec.	2.71
Muon momentum resolution (MS)	0.06
Muon momentum resolution (ID)	0.10
Muon momentum scale	0.98
$E_T^{\text{miss}}$ Soft Jet scale	0.79
$E_T^{\text{miss}}$ Soft Jet resolution	0.92
Luminosity	3.60
DD QCD norm.	0.49
All syst.	40.72
Stat. MC	1.82
Stat. Data	1.66

# Chapter 8

## Summary

As mentioned in the introduction (Chapter 1), the top quark has an interesting position in the Standard Model, due to its high mass compared to the other quarks and its small lifetime. It allows us to study a “bare quark” decay, while other quarks hadronise. It has a strong coupling to the Higgs boson and the top-antitop production is a main background in many analyses, such as the top-antitop-Higgs search. Furthermore, in many hypotheses Beyond the Standard Model it is expected that the top quark plays an important role with particles which were not identified so far.

The ATLAS detector (Chapter 3) is an excellent environment to study the top quark physics in detail. It is particularly important to detect  $b$ -jets in the detector, since the top quark decays 99% of the times to a  $W$ -boson and a  $b$ -quark. A performance study of the  $b$ -jet trigger in ATLAS was done in Chapter 4, showing that the detector is ready to trigger on  $b$ -jets.

Aiming at a study of the Standard Model top quark production, a measurement of the top-antitop pair production cross section times the branching ratio in which there is an electron or a muon in the final state was measured in data as a function of the jet multiplicity in Chapter 5. An unfolding procedure was devised to correct for the detector effects, with a systematic uncertainty estimate for the corrections. The jet multiplicity in the final state shows the effect of the QCD radiation in each bin. A comparison of Monte Carlo generators, including the matrix element generators and the parton shower simulation variations, was done in the final unfolded result. It shows that MC@NLO and Alpgen+Pythia with the  $\alpha_S$  increased variation describe the data very badly, but Alpgen+Herwig, Powheg and Alpgen+Pythia with the  $\alpha_S$  decreased variation describe the data better.

Although the Standard Model results can be measured directly using the

top-antitop cross section analysis, another way of testing the Standard Model prediction is to test the hypothesis that other models describe data better. A few models expect unverified particles to decay into top-antitop pairs and, in Chapters 6 and 7, they were tested with ATLAS data. The hypothesis that they are valid was excluded with 95% Confidence Level for  $m_{t\bar{t}} \in [0.5 \text{ TeV}, 1.8 \text{ TeV}]$  for a Topcolor  $Z'$  model and  $m_{t\bar{t}} \in [0.5 \text{ TeV}, 2.07 \text{ TeV}]$  for the Kaluza-Klein gluons.

The analyses in the thesis show that, although there are still open points in the theory, the Standard Model describes the top-antitop production better than the studied alternatives. It also has a good description of the top-antitop jet multiplicity for some Monte Carlo generators, while others need improvement.

# Appendix A

## Top-antitop + jets control plot distributions

Kinematic distributions for reconstructed objects and their comparison to data driven background and MC predictions in the  $t\bar{t}$  + jets cross section measurement, before the unfolding procedure, using the Alpgen+Herwig  $t\bar{t}$  MC sample, are shown in Figures A.1, A.2, A.3, A.4, A.5, A.6, A.7, A.8 for the electron and muon channels. The background was estimated using simulation and data-driven techniques, as described in Sections 5.2, 5.4 and 5.5. The event selection follows the procedure described in Section 5.3.



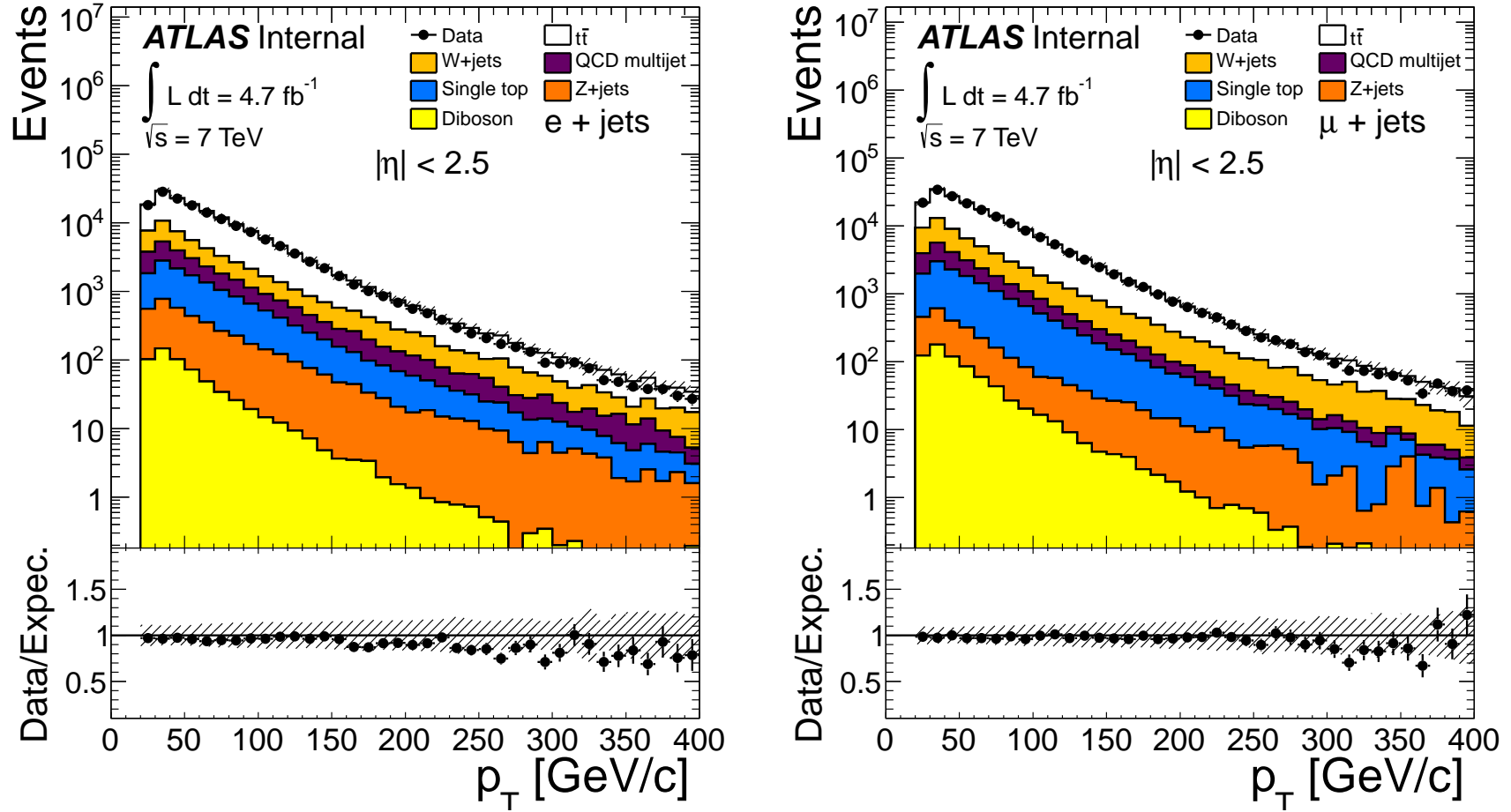


Figure A.1: Data to expected signal and background comparison of all jets  $p_T$  from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

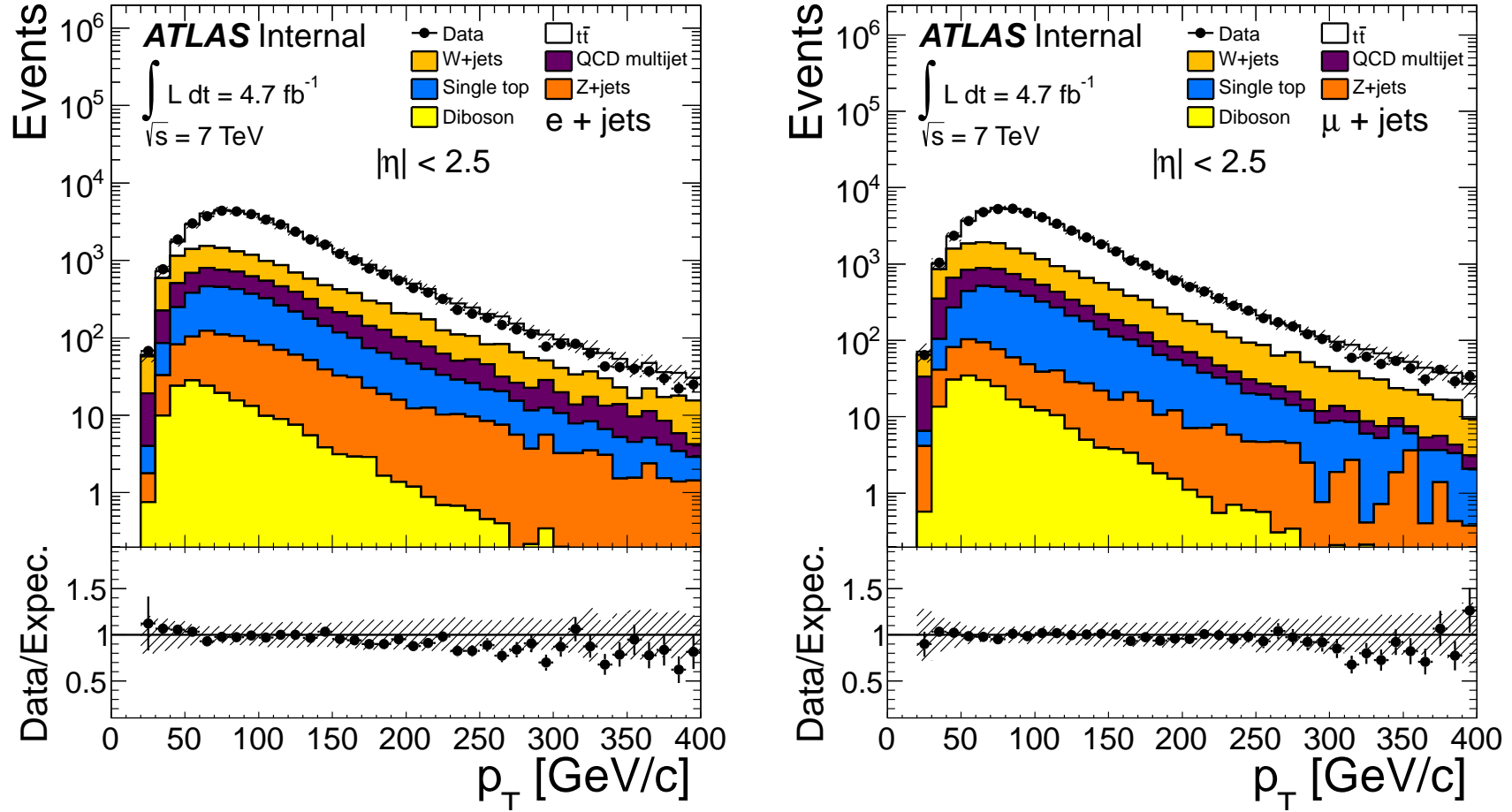


Figure A.2: Data to expected signal and background comparison of the highest transverse momentum jet  $p_T$  from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

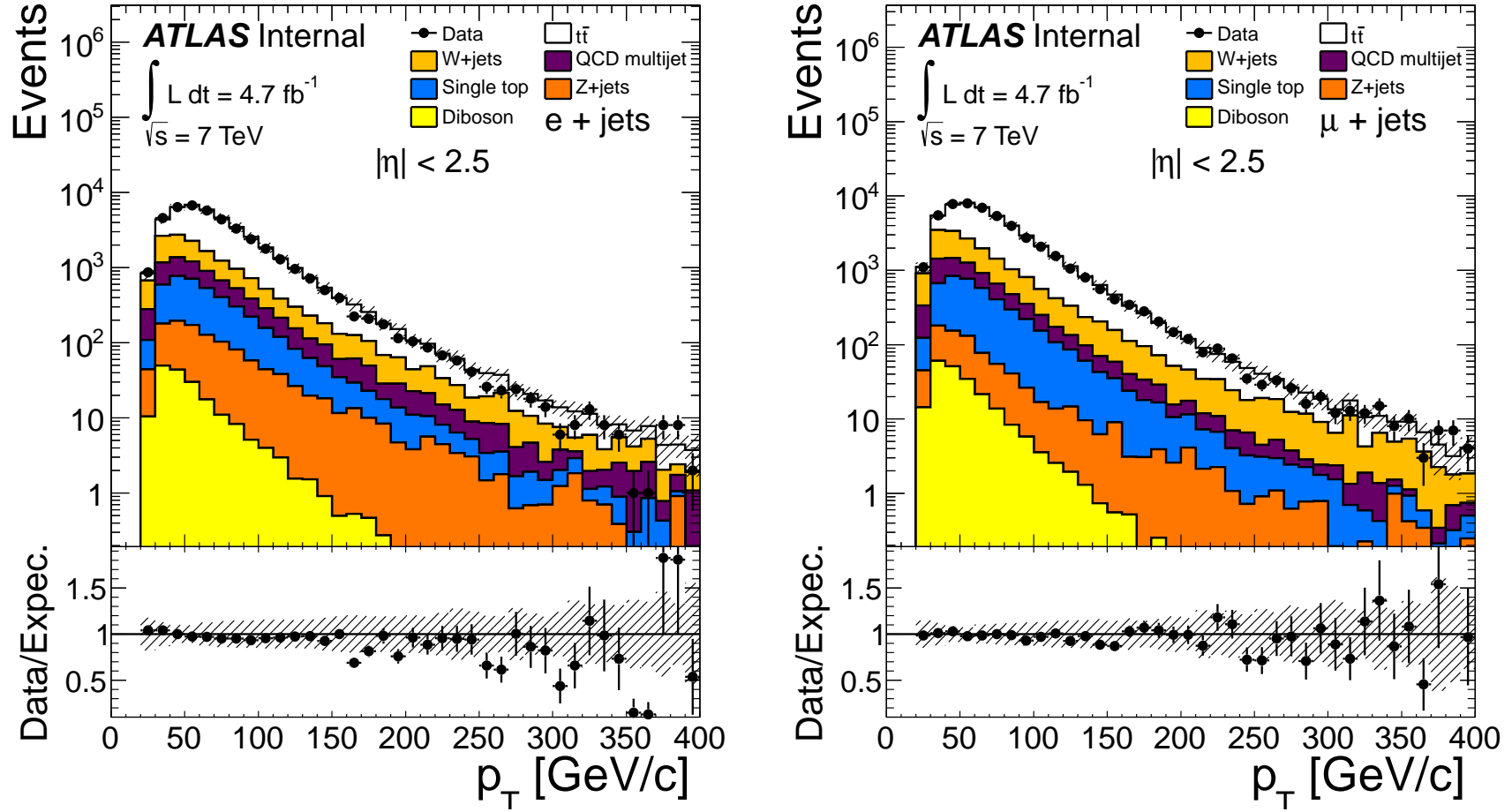


Figure A.3: Data to expected signal and background comparison of the second highest transverse momentum jet  $p_T$  from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

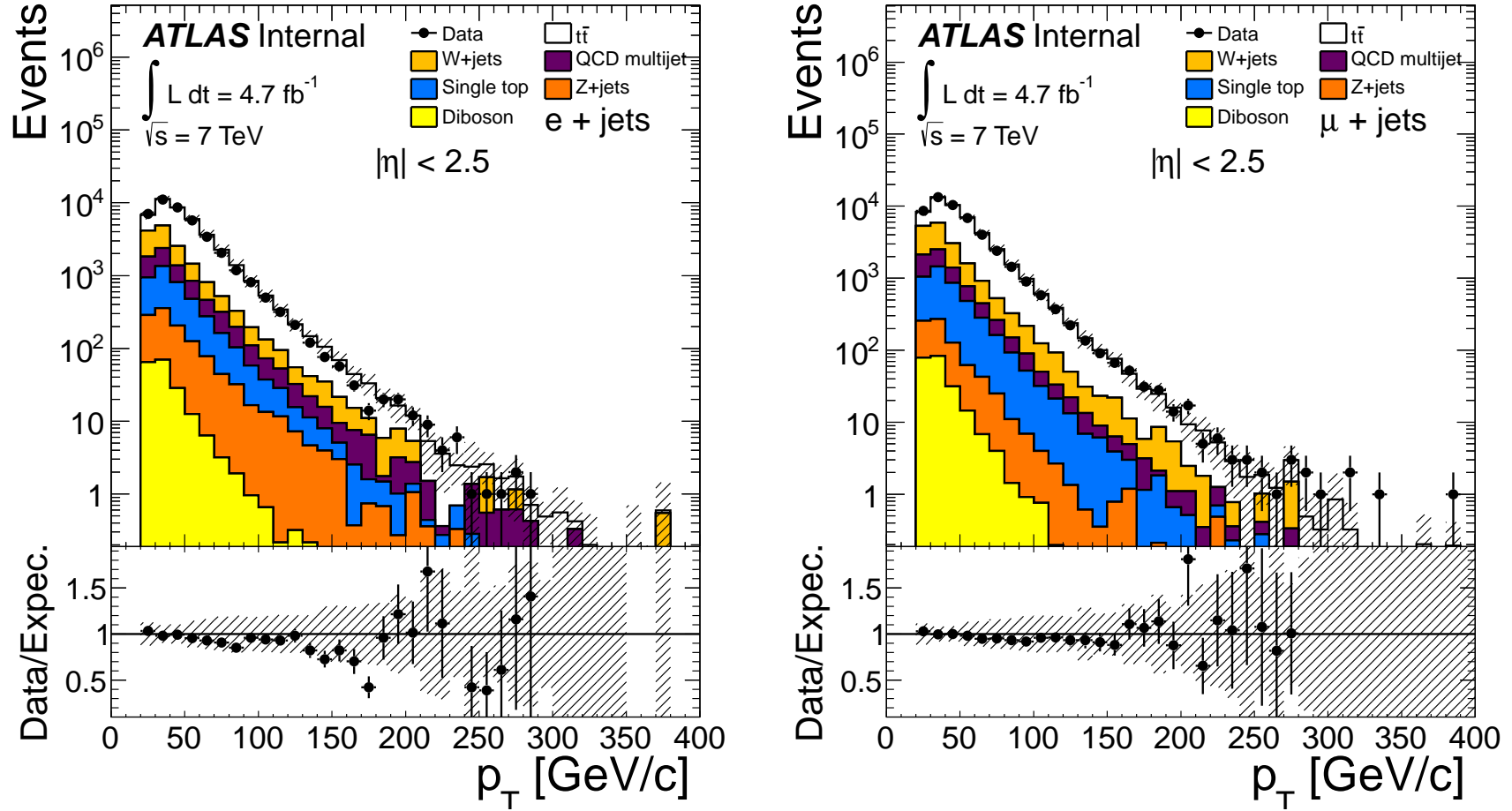


Figure A.4: Data to expected signal and background comparison of the third highest transverse momentum jet  $p_T$  from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

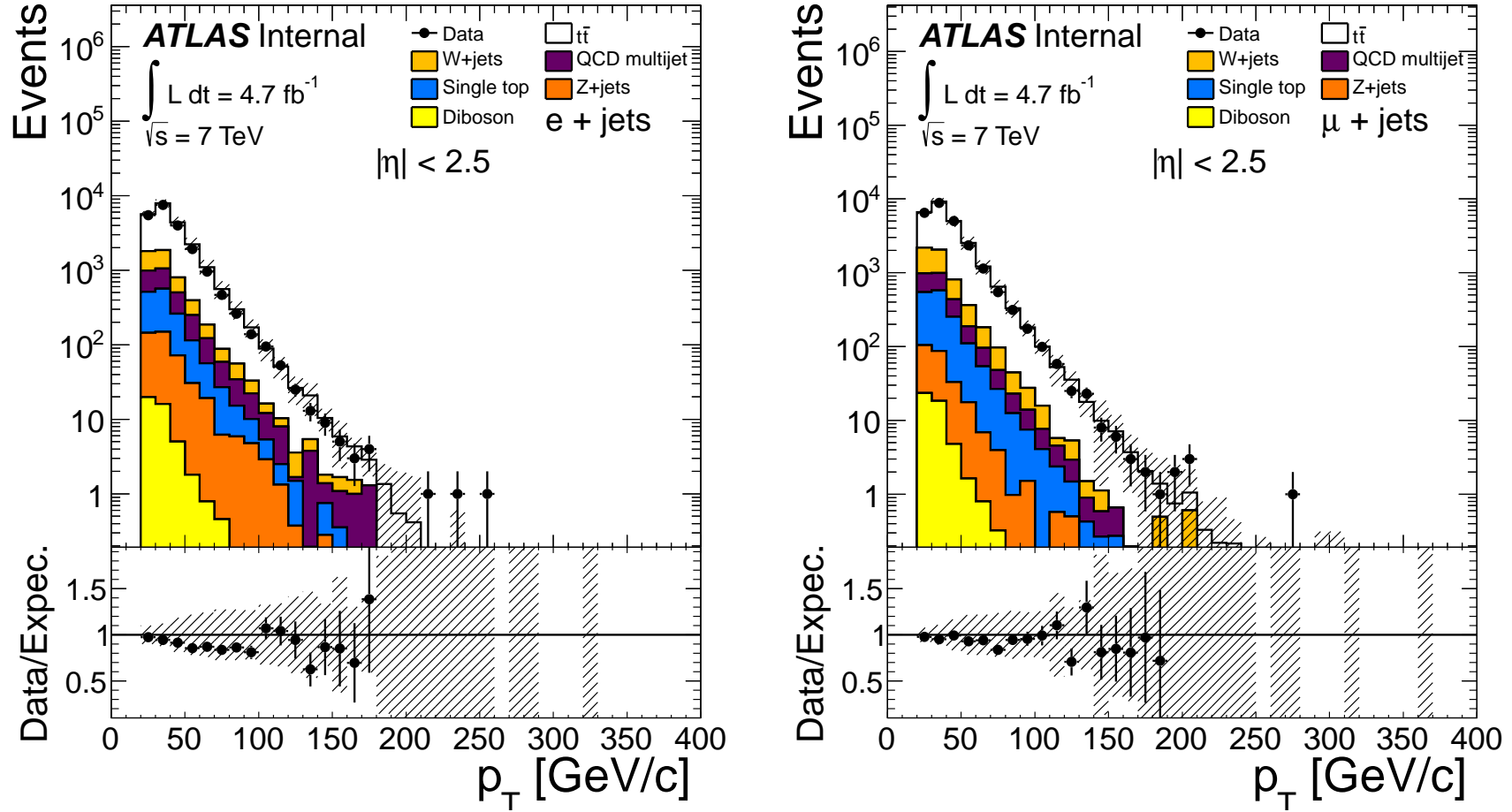


Figure A.5: Data to expected signal and background comparison of the fourth highest transverse momentum jet  $p_T$  from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

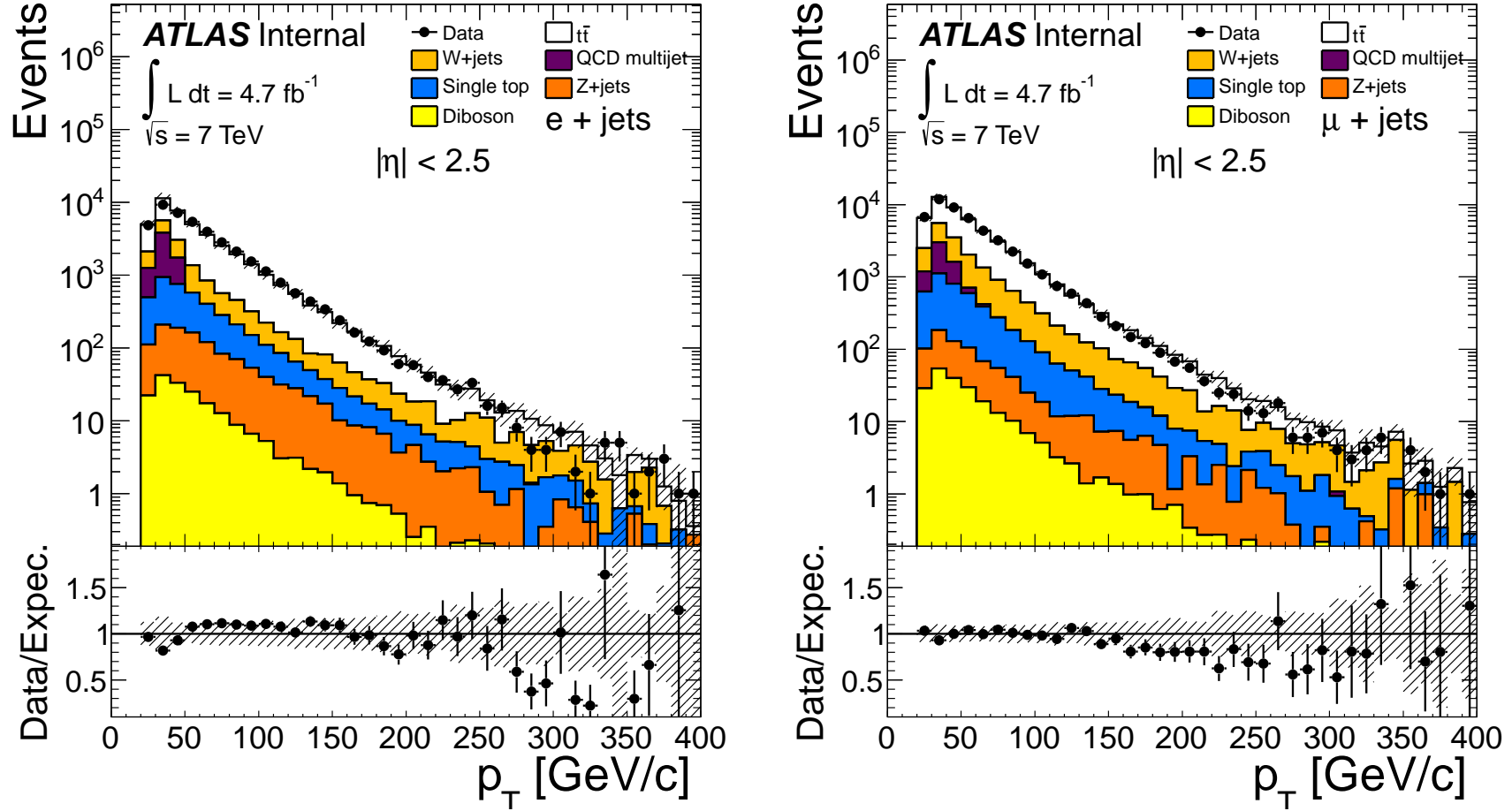


Figure A.6: Data to expected signal and background comparison of the lepton transverse momentum from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

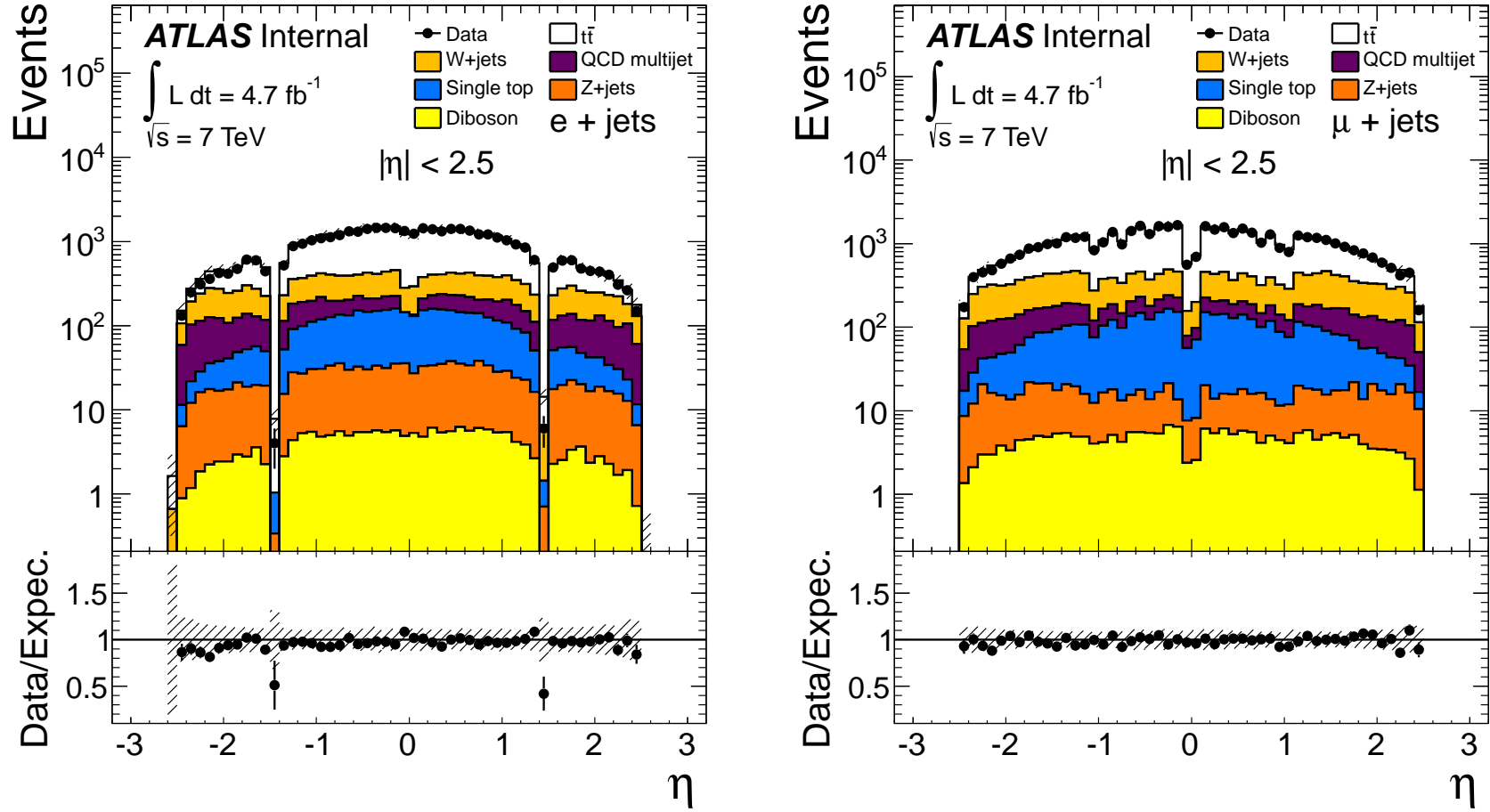


Figure A.7: Data to expected signal and background comparison of the lepton pseudo-rapidity from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.

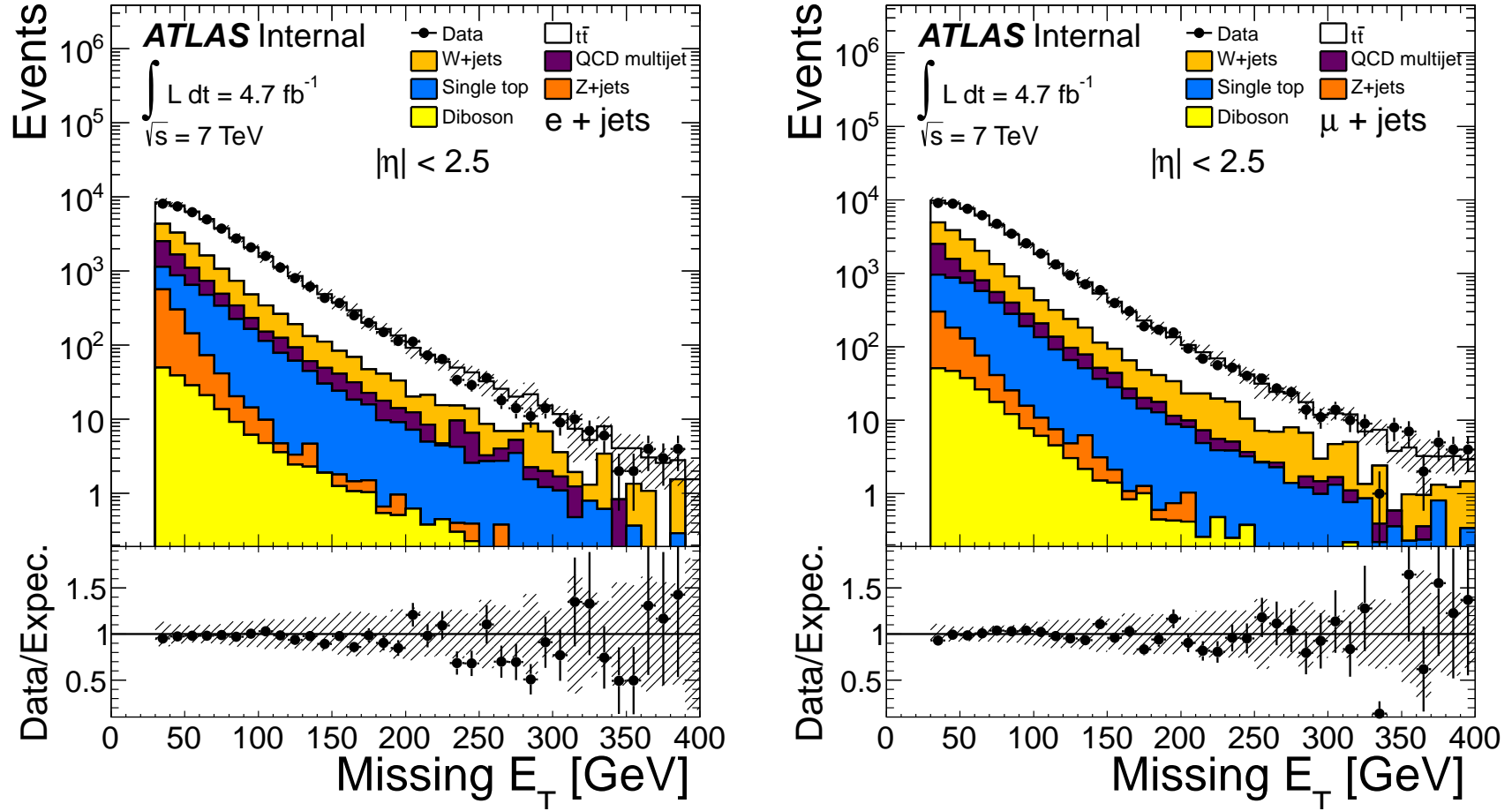


Figure A.8: Data to expected signal and background comparison of the missing transverse energy from reconstructed objects using the electron (left) and muon (right) channels for the event selection in the top-antitop jet multiplicity analysis with a minimum jet transverse momentum of 25 GeV. The Alpgen+Herwig [44, 48, 49]  $t\bar{t}$  MC sample was used within the data driven and MC predictions.



# Bibliography

- [1] C.C.W. Taylor. *The Atomists: Leucippus and Democritus. Fragments, A Text and Translation with Commentary*. University of Toronto Press, Toronto, 1999.
- [2] Mrinalkanti Gangopadhyaya. *Indian Atomism: History and Sources*. Humanities Press, New Jersey, 1981.
- [3] Claude Cohen-Tannoudji, Bernard Diu, and Frank Laló. *Quantum Mechanics*. Wiley VCH, 2006.
- [4] Michael E. Peskin and Dan V. Schroeder. *An Introduction To Quantum Field Theory (Frontiers in Physics)*. Westview Press, 1995.
- [5] CDF Collaboration. Observation of Top Quark Production in  $p\bar{p}$  Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74(14):2626–2631, Apr 1995.
- [6] S. Abachi et al. (D0 Collaboration). Observation of the top quark. *Phys. Rev. Lett.*, 74(14):2632–2637, Apr 1995.
- [7] J. Beringer et al. Review of Particle Physics (RPP). *Phys. Rev.*, D86:010001, 2012.
- [8] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012, arXiv: hep-ex/1207.7214.
- [9] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012, arXiv: hep-ex/1207.7235.
- [10] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys.Lett.*, 12:132–133, 1964.

- [11] ATLAS Collaboration, G. Aad, and et al. The ATLAS Experiment at the CERN Large Hadron Collider. In *ATLAS Detector Status and Physics Startup Plans*, 2008.
- [12] Oliver Sim Brning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004.
- [13] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.
- [14] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [15] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.
- [16] Howard Georgi and Sheldon L. Glashow. Unified weak and electromagnetic interactions without neutral currents. *Phys. Rev. Lett.*, 28:1494–1497, May 1972.
- [17] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.
- [18] H David Politzer. Asymptotic freedom: An approach to strong interactions. *Physics Reports*, 14(4):129 – 180, 1974.
- [19] David J. Gross and Frank Wilczek. Asymptotically free gauge theories. i. *Phys. Rev. D*, 8:3633–3652, Nov 1973.
- [20] Steven Weinberg. The Making of the standard model. *Eur.Phys.J.*, C34:5–13, 2004, arXiv: hep-ph/0401010.
- [21] G.’t Hooft. Renormalizable Lagrangians for massive Yang-Mills fields. *Nuclear Physics B*, 35(1):167 – 188, 1971.
- [22] G. ’t Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44(1):189 – 213, 1972.
- [23] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [24] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

- [25] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.
- [26] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973.
- [27] S. Fukuda et al. Determination of solar neutrino oscillation parameters using 1496 days of Super-Kamiokande I data. *Phys.Lett.*, B539:179–187, 2002, arXiv: hep-ex/0205075.
- [28] Q.R. Ahmad et al. Measurement of the rate of  $\nu/e + d \rightarrow p + p + e^-$  interactions produced by B-8 solar neutrinos at the Sudbury Neutrino Observatory. *Phys.Rev.Lett.*, 87:071301, 2001, arXiv: nucl-ex/0106015.
- [29] Q.R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys.Rev.Lett.*, 89:011301, 2002, arXiv: nucl-ex/0204008.
- [30] Joseph R. Incandela, Arnulf Quadt, Wolfgang Wagner, and Daniel Wicke. Status and Prospects of Top-Quark Physics. *Prog. Part. Nucl. Phys.*, 63:239–292, 2009, arXiv: hep-ex/0904.2499.
- [31] Wolfgang Wagner. Top quark physics in hadron collisions. *Rept.Prog.Phys.*, 68:2409–2494, 2005, arXiv: hep-ph/0507207.
- [32] Rajan Gupta. Introduction to lattice QCD: Course. pages 83–219, 1997, arXiv: hep-lat/9807028.
- [33] Paul Langacker. Structure of the standard model. 1995, arXiv: hep-ph/0304186.
- [34] Georges Aad et al. Search for FCNC single top-quark production at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *Phys.Lett.*, B712:351–369, 2012, arXiv: hep-ex/1203.0529.
- [35] A. Quadt. Top quark physics at hadron colliders. *European Physical Journal C*, 48:835–1000, December 2006.
- [36] E. Shabalina. Top Quark Production Cross Section at the Tevatron. 2006, arXiv: hep-ex/0605045.

- [37] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Hoche, et al. General-purpose event generators for LHC physics. *Phys.Rept.*, 504:145–233, 2011, arXiv: hep-ph/1101.2599.
- [38] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 0207:012, 2002, arXiv: hep-ph/0201195.
- [39] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010, arXiv: hep-ph/1007.2241.
- [40] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, et al. Parton distributions with LHC data. *Nucl.Phys.*, B867:244–289, 2013, arXiv: hep-ph/1207.1303.
- [41] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009, arXiv: hep-ph/0901.0002.
- [42] T. Stelzer and W.F. Long. Automatic generation of tree level helicity amplitudes. *Comput.Phys.Commun.*, 81:357–371, 1994, arXiv: hep-ph/9401258.
- [43] Borut Paul Kersevan and Elzbieta Richter-Was. The Monte Carlo event generator AcerMC version 1.0 with interfaces to PYTHIA 6.2 and HERWIG 6.3. *Comput.Phys.Commun.*, 149:142–194, 2003, arXiv: hep-ph/0201302.
- [44] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 0307:001, 2003, arXiv: hep-ph/0206293.
- [45] Francesco Caravaglios and Mauro Moretti. An algorithm to compute Born scattering amplitudes without Feynman graphs. *Phys.Lett.*, B358:332–338, 1995, arXiv: hep-ph/9507237.
- [46] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977.

- [47] G. Dissertori, I.G. Knowles, and M. Schmelling. *Quantum Chromodynamics: High Energy Experiments and Theory*. International Series of Monographs on Physics. OUP Oxford, 2009.
- [48] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001, arXiv: hep-ph/0011363.
- [49] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6.5 release note. 2002, arXiv: hep-ph/0210213.
- [50] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006, arXiv: hep-ph/0603175.
- [51] S. Catani, F. Krauss, R. Kuhn, and B.R. Webber. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001, arXiv: hep-ph/0109231.
- [52] Johan Alwall, Stefan Hoche, F. Krauss, N. Lavesson, L. Lonnblad, et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur.Phys.J.*, C53:473–500, 2008, arXiv: hep-ph/0706.2569.
- [53] Roberto Chierici. Monte Carlo generators for the LHC. *DESY-PROC-2009-06*, pages 32–41, 2009. Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI’08).
- [54] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 0711:070, 2007, arXiv: hep-ph/0709.2092.
- [55] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002, arXiv: hep-ph/0204244.
- [56] Stefano Frixione and Bryan R. Webber. The MC and NLO 3.4 Event Generator. 2008, arXiv: hep-ph/0812.0770.
- [57] Stefano Frixione, Fabian Stoeckli, Paolo Torrielli, Bryan R. Webber, and Chris D. White. The MC@NLO 4.0 Event Generator. 2010, arXiv: hep-ph/1010.0819.

- [58] Bobby Samir Acharya, Francesca Cavallari, Gennaro Corcella, Riccardo Di Sipio, and Giovanni Petrucciani. Monte Carlo generators for top quark physics at the LHC. *Nuovo Cim.*, B123:415–420, 2008, arXiv: hep-ph/0804.4122.
- [59] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2 - 3):31 – 145, 1983.
- [60] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour, and L. Stanco. HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. *Computer Physics Communications*, 67(3):465 – 508, 1992.
- [61] S. Agostinelli et al. Geant4 - a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.
- [62] P. Bartalini, E.L. Berger, B. Blok, G. Calucci, R. Corke, et al. Multi-Parton Interactions at the LHC. 2011, arXiv: hep-ph/1111.0469.
- [63] J.M. Butterworth, Jeffrey R. Forshaw, and M.H. Seymour. Multiparton interactions in photoproduction at HERA. *Z.Phys.*, C72:637–646, 1996, arXiv: hep-ph/9601371.
- [64] R.D. Peccei, S. Peris, and X. Zhang. Non-standard couplings of the top quark and precision measurements of the electroweak theory. *Nuclear Physics B*, 349(2):305 – 322, 1991.
- [65] Stephen P. Martin. A Supersymmetry primer. 1997, arXiv: hep-ph/9709356.
- [66] Christopher T. Hill. Topcolor assisted technicolor. *Phys.Lett.*, B345:483–489, 1995, arXiv: hep-ph/9411426.
- [67] Robert M. Harris, Christopher T. Hill, and Stephen J. Parke. Cross-section for topcolor Z-prime(t) decaying to t anti-t. 1999, arXiv: hep-ph/9911288.
- [68] Robert M. Harris and Supriya Jain. Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop. *Eur.Phys.J.*, C72:2072, 2012, arXiv: hep-ph/1112.4928.

- [69] Ben Lillie, Lisa Randall, and Lian-Tao Wang. The Bulk RS KK-gluon at the LHC. *JHEP*, 0709:074, 2007, arXiv: hep-ph/0701166.
- [70] Ben Lillie, Jing Shu, and Timothy M.P. Tait. Kaluza-Klein Gluons as a Diagnostic of Warped Models. *Phys.Rev.*, D76:115016, 2007, arXiv: hep-ph/0706.3960.
- [71] ATLAS Collaboration. *ATLAS Detector and Physics Performance Technical Design Report*. CERN, 1999.
- [72] ATLAS Collaboration. *ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report*. CERN, 1999.
- [73] ATLAS Collaboration. *ATLAS Inner Detector Technical Design Report*. CERN, 1997.
- [74] ATLAS Collaboration. *Liquid Argon Calorimeter Technical Design Report*. CERN, 1996.
- [75] ATLAS Collaboration. *Tile Calorimeter Technical Design Report*. CERN, 1996.
- [76] Georges Aad et al. Performance of the ATLAS Trigger System in 2010. *Eur.Phys.J.*, C72:1849, 2012, arXiv: hep-ex/1110.1530.
- [77] Per Ola Hansson Adrian. The ATLAS b-Jet Trigger. 2011, arXiv: hep-ex/1111.4190.
- [78] Georges Aad et al. Improved luminosity determination in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC. *Eur.Phys.J.*, C73:2518, 2013, arXiv: hep-ex/1302.4393.
- [79] Georges Aad et al. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur.Phys.J.*, C72:1909, 2012, arXiv: hep-ex/1110.3174.
- [80] Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector. Technical Report ATLAS-CONF-2011-063, CERN, Geneva, Apr 2011.
- [81] Muon Momentum Resolution in First Pass Reconstruction of  $pp$  Collision Data Recorded by ATLAS in 2010. Technical Report ATLAS-CONF-2011-046, CERN, Geneva, Mar 2011.

- [82] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008, arXiv: hep-ph/0802.1189.
- [83] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys.Rev.*, D48:3160–3166, 1993, arXiv: hep-ph/9305266.
- [84] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012, arXiv: hep-ph/1111.6097.
- [85] Georges Aad et al. Jet energy measurement with the ATLAS detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV. *Eur.Phys.J.*, C73:2304, 2013, arXiv: hep-ex/1112.6426.
- [86] C. Cojocaru et al. Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region in beam tests. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 531(3):481 – 514, 2004.
- [87] Jet energy scale and its systematic uncertainty in proton-proton collisions at  $\sqrt{s} = 7$  TeV with ATLAS 2011 data. Technical Report ATLAS-CONF-2013-004, CERN, Geneva, Jan 2013.
- [88] Performance of large-R jets and jet substructure reconstruction with the ATLAS detector. Technical Report ATLAS-CONF-2012-065, CERN, Geneva, Jul 2012.
- [89] Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy  $\sqrt{s} = 7$  TeV. Technical Report ATLAS-CONF-2010-054, CERN, Geneva, Jul 2010.
- [90] Measuring the b-tag efficiency in a top-pair sample with  $4.7 \text{ fb}^{-1}$  of data from the ATLAS detector. Technical Report ATLAS-CONF-2012-097, CERN, Geneva, Jul 2012.
- [91] Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.



- [92] Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with  $5 \text{ fb}^{-1}$  of Data from the ATLAS Detector. Technical Report ATLAS-CONF-2012-043, CERN, Geneva, Mar 2012.
- [93] Georges Aad et al. Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS. *Eur.Phys.J.*, C72:1844, 2012, arXiv: hep-ex/1108.5602.
- [94] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuszi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur.Phys.J.*, C71:1753, 2011, arXiv: hep-ph/1107.5909.
- [95] Measurement of the jet multiplicity in top anti-top final states produced in 7 TeV proton-proton collisions with the ATLAS detector. Technical Report ATLAS-CONF-2012-155, CERN, Geneva, Nov 2012.
- [96] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys.Rev.*, D82:074018, 2010, arXiv: hep-ph/1005.3457.
- [97] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, et al. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput.Phys.Commun.*, 182:1034–1046, 2011, arXiv: hep-ph/1007.1327.
- [98] Nikolaos Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys.Rev.*, D83:091503, 2011, arXiv: hep-ph/1103.2792.
- [99] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a  $W$  or  $H$ . *Phys.Rev.*, D82:054018, 2010, arXiv: hep-ph/1005.4451.
- [100] Nikolaos Kidonakis. NNLL resummation for s-channel single top quark production. *Phys.Rev.*, D81:054028, 2010, arXiv: hep-ph/1001.5034.
- [101] Georges Aad et al. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$ . *Eur.Phys.J.*, C71:1577, 2011, arXiv: hep-ex/1012.1792.
- [102] Sergey Alekhin, Simone Alioli, Richard D. Ball, Valerio Bertone, Johannes Blumlein, et al. The PDF4LHC Working Group Interim Report. 2011, arXiv: hep-ph/1101.0536.

- [103] Luminosity Determination in pp Collisions at  $\sqrt{s} = 7$  TeV using the ATLAS Detector in 2011. Technical Report ATLAS-CONF-2011-116, CERN, Geneva, Aug 2011.
- [104] Georges Aad et al. Measurement of  $t\bar{t}$  production with a veto on additional central jet activity in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector. *Eur.Phys.J.*, C72:2043, 2012, arXiv: hep-ex/1203.5015.
- [105] Georges Aad et al. A search for  $t\bar{t}$  resonances in the lepton plus jets final state with ATLAS using 4.7 fb<sup>-1</sup> of pp collisions at  $\sqrt{s} = 7$  TeV. *Phys.Rev.*, D88:012004, 2013, arXiv: hep-ex/1305.2756.
- [106] A search for ttbar resonances in the lepton plus jets final state using 4.66 fb<sup>-1</sup> of pp collisions at  $\sqrt{s} = 7$  TeV. Technical Report ATLAS-CONF-2012-136, CERN, Geneva, Sep 2012.
- [107] Martin Beneke, Michal Czakon, Pietro Falgari, Alexander Mitov, and Christian Schwinn. Threshold expansion of the  $gg(q\bar{q}) \rightarrow Q\bar{Q} + X$  cross section at O( $\alpha_s^4$ ). *Phys.Lett.*, B690:483–490, 2010, arXiv: hep-ph/0911.5166.
- [108] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Uncertainties on  $\alpha_s$  in global PDF analyses and implications for predicted hadronic cross sections. *Eur.Phys.J.*, C64:653–680, 2009, arXiv: hep-ph/0905.3531.
- [109] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys.Lett.*, B710:612–622, 2012, arXiv: hep-ph/1111.5869.
- [110] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. 2011, arXiv: hep-ph/1112.5675.
- [111] Georges Aad et al. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at  $\sqrt{s} = 7$  TeV. *Eur.Phys.J.*, C71:1577, 2011, arXiv: hep-ex/1012.1792.
- [112] Aneesh V. Manohar and Michael Trott. Electroweak Sudakov Corrections and the Top Quark Forward-Backward Asymmetry. *Phys.Lett.*, B711:313–316, 2012, arXiv: hep-ph/1201.3926.

- [113] G. Choudalakis. On hypothesis testing, trials factor, hypertexts and the BumpHunter. January 2011, arXiv: physics.data-an/1101.0390.
- [114] I. Bertram et al. A Recipe for the construction of confidence limits. Technical Report FERMILAB-TM-2104, D0-NOTE-3476, D0-NOTE-2775-A, FERMILAB, 2000.
- [115] A search for  $t\bar{t}$  resonances in the lepton plus jets final state with ATLAS using 14 fb<sup>1</sup> of pp collisions at  $\sqrt{s} = 8$  TeV. Technical Report ATLAS-CONF-2013-052, CERN, Geneva, May 2013.
- [116] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 1002:084, 2010, arXiv: hep-ph/0912.1342.